

**22nd International Conference in Central Europe
on
Computer Graphics, Visualization and Computer Vision**

in co-operation with

EUROGRAPHICS Association

WSCG 2014

Poster Papers Proceedings

Edited by

Vaclav Skala, University of West Bohemia, Czech Republic

**22nd International Conference in Central Europe
on
Computer Graphics, Visualization and Computer Vision**

in co-operation with

EUROGRAPHICS Association

WSCG 2014

Poster Papers Proceedings

Edited by

Vaclav Skala, University of West Bohemia, Czech Republic

Vaclav Skala – Union Agency

WSCG 2014 – Poster Papers Proceedings

Editor: Vaclav Skala
c/o University of West Bohemia, Univerzitni 8
CZ 306 14 Plzen
Czech Republic
skala@kiv.zcu.cz <http://www.VaclavSkala.eu>

Managing Editor: Vaclav Skala

Published and printed by:
Vaclav Skala – Union Agency
Na Mazinách 9
CZ 322 00 Plzen
Czech Republic <http://www.UnionAgency.eu>

Hardcopy: *ISBN 978-80-86943-72-5*

WSCG 2014

International Program Committee

Balcisoy,Selim (Turkey)	Oliveira,Manuel M. (Brazil)
Benes,Bedrich (United States)	Oyarzun Laura,Cristina (Germany)
Benger,Werner (United States)	Pan,Rongjiang (China)
Bittner,Jiri (Czech Republic)	Patow,Gustavo (Spain)
Daniel,Marc (France)	Platis,Nikos (Greece)
Daniels,Karen (United States)	Puppo,Enrico (Italy)
de Geus,Klaus (Brazil)	Renaud,Christophe (France)
Debelov,Victor (Russia)	Richardson,John (United States)
Feito,Francisco (Spain)	Rojas-Sola,Jose Ignacio (Spain)
Ferguson,Stuart (United Kingdom)	Sabine,Coquillart (France)
Flaquer,Juan (Spain)	Santos,Luis Paulo (Portugal)
Gain,James (South Africa)	Segura,Rafael (Spain)
Gavrilova,M. (Canada)	Semwal,Sudhanshu (United States)
Giannini,Franca (Italy)	Sousa,A.Augusto (Portugal)
Guthe,Michael (Germany)	Stroud,Ian (Switzerland)
Herout,Adam (Czech Republic)	Szecs,Laszlo (Hungary)
Choi,Sunghee (Korea)	Teschner,Matthias (Germany)
Juan,M.-Carmen (Spain)	Theussl,Thomas (Saudi Arabia)
Kalra,Prem K. (India)	Tokuta,Alade (United States)
Kellomaki,Timo (Finland)	Wuensche,Burkhard,C. (New Zealand)
Kim,HyungSeok (Korea)	Wuethrich,Charles (Germany)
Klosowski,James (United States)	Zemcik,Pavel (Czech Republic)
Kurt,Murat (Turkey)	Zitova,Barbara (Czech Republic)
Max,Nelson (United States)	Zwettler,Gerald (Austria)
Muller,Heinrich (Germany)	
Murtagh,Fionn (United Kingdom)	

WSCG 2014

Board of Reviewers

Abad, Francisco	Ferguson, Stuart	Jeschke, Stefan
Adzhiev, Valery	Fernandes, Antonio	Joan-Arinyo, Robert
Agathos, Alexander	Flaquer, Juan	Jones, Mark
Alvarado, Adriana	Fuenfzig, Christoph	Juan, M.-Carmen
Assarsson, Ulf	Gain, James	Kalra, Prem K.
Ayala, Dolors	Galo, Mauricio	Kämpe, Viktor
Backfrieder, Werner	Garcia Hernandez, Ruben Jesus	Kanai, Takashi
Balcisoy, Selim	Garcia-Alonso, Alejandro	Kellomaki, Timo
Barbosa, Joao	Gavrilova, M.	Kim, HyungSeok
Barthe, Loic	Geus, Klaus de	Klosowski, James
Battiato, Sebastiano	Giannini, Franca	Kolcun, Alexej
Benes, Bedrich	Gobron, Stephane	Krivanek, Jaroslav
Benger, Werner	Gobron, Stephane	Kurillo, Gregorij
Bilbao, Javier, J.	Gois, Joao Paulo	Kurt, Murat
Billeter, Markus	Grau, Sergi	Kyratzi, Sofia
Biri, Venceslas	Gudukbay, Ugur	Larboulette, Caroline
Birra, Fernando	Guthe, Michael	Lee, Jong Kwan
Bittner, Jiri	Haberdar, Hakan	Liu, Damon Shing-Min
Bosch, Carles	Hansford, Dianne	Lopes, Adriano
Bourdin, Jean-Jacques	Haro, Antonio	Loscos, Celine
Brun, Anders	Hasler, Nils	Lutteroth, Christof
Bruni, Vittoria	Hast, Anders	Maciel, Anderson
Buehler, Katja	Hernandez, Benjamin	Mandl, Thomas
Bulo, Samuel Rota	Herout, Adam	Manzke, Michael
Cakmak, Hueseyin Kemal	Herrera, Tomas Lay	Marras, Stefano
Camahort, Emilio	Hicks, Yulia	Masia, Belen
Casciola, Giulio	Hildenbrand, Dietmar	Masood, Syed Zain
Cline, David	Hinkenjann, Andre	Masso, Jose Pascual Molina
Coquillart, Sabine	Horain, Patrick	Matey, Luis
Cosker, Darren	Chaine, Raphaelle	Max, Nelson
Daniel, Marc	Choi, Sunghee	Melendez, Francho
Daniels, Karen	Chover, Miguel	Meng, Weiliang
Debelov, Victor	Chrysanthou, Yiorgos	Mestre, Daniel, R.
Drechsler, Klaus	Chuang, Yung-Yu	Meyer, Alexandre
Durikovic, Roman	Iglesias, Jose A.	Molla, Ramon
Eisemann, Martin	Ihrke, Ivo	Montrucchio, Bartolomeo
Erbacher, Robert	Iwasaki, Kei	Morigi, Serena
Feito, Francisco	Jato, Oliver	Mukai, Tomohiko

Muller, Heinrich
Munoz, Adolfo
Murtagh, Fionn
Okabe, Makoto
Oliveira, Joao
Oliveira, Manuel M.
Oyarzun, Cristina Laura
Pan, Rongjiang
Papaioannou, Georgios
Paquette, Eric
Pasko, Galina
Patane, Giuseppe
Patow, Gustavo
Pedrini, Helio
Pereira, Joao Madeiras
Peters, Jorg
Pina, Jose Luis
Platis, Nikos
Post, Frits, H.
Puig, Anna
Puppo, Enrico
Puppo, Enrico
Rafferty, Karen
Raffin, Romain
Renaud, Christophe
Reshetouski, Ilya
Reshetov, Alexander
Ribardiere, Mickael
Ribeiro, Roberto
Richardson, John

Ritter, Marcel
Rojas-Sola, Jose Ignacio
Rokita, Przemyslaw
Rudomin, Isaac
Sacco, Marco
Sadlo, Filip
Salvetti, Ovidio
Sanna, Andrea
Santos, Luis Paulo
Sapidis, Nickolas, S.
Savchenko, Vladimir
Segura, Rafael
Seipel, Stefan
Sellent, Anita
Semwal, Sudhanshu
Sheng, Bin
Sheng, Yu
Shesh, Amit
Schmidt, Johanna
Sik-Lanyi, Cecilia
Sintorn, Erik
Sirakov, Nikolay Metodiev
Sourin, Alexei
Sousa, A. Augusto
Sramek, Milos
Stroud, Ian
Subsol, Gerard
Sundstedt, Veronica
Svoboda, Tomas
Szecsi, Laszlo

Tang, Min
Teschner, Matthias
Theussl, Thomas
Tian, Feng
Tokuta, Alade
Torrens, Francisco
Trapp, Matthias
Tytkowski, Krzysztof
Umlauf, Georg
Vergeest, Joris
Vitulano, Domenico
Vosinakis, Spyros
Walczak, Krzysztof
Wan, Liang
Wu, Shin-Ting
Wuensche, Burkhard, C.
Wuethrich, Charles
Xin, Shi-Qing
Xu, Fei Dai Dongrong
Yoshizawa, Shin
Yue, Yonghao
Yue, Yonghao
Zalik, Borut
Zemcik, Pavel
Zhang, Xinyu
Zhao, Qiang
Zheng, Youyi
Zitova, Barbara
Zwettler, Gerald

WSCG 2014

Poster Proceedings

Contents

	Page
Tian,L.: Ocean wave simulation by the mix of FFT and Perlin Noise	1
Antemijczuk,P., Magiera,M., Lehmann,S., Cuttone,A., Larsen,J.: Visualizing multi-channel networks	5
Benziane,S., Merouane,A.,Meroune,A.: Parallel extraction hand vein biometric parameter's using a low cost IR Imaging system	9
Bezuglov,D.A., Bezuglov,Yu.D., Shvidchenko,S.A.: Method of Discrete Wavelet Analysis of Edges on the Random Background	15
Gdawiec,K., Kotarski,W., Lisowska,A.: Polynomiography with Non-standard Iterations	21
Lahdenoja,O., Sääntti,T., Laiho,M., Poikonen,J.: Spatter Tracking in Laser- and Manual Arc Welding with Sensor-level Pre-processing	27
Milet,T., Kobrtek,J., Zemčík,P., Pečiva,J.: Fast and Robust Tessellation-Based Silhouette Shadows	33
Akagi,Y., Kawasaki,H.: A method of micro facial expression recognition based on dense facial motion data	39
Endler,M., Lobachev,O., Guthe,M.: Human-Computer Interaction Using Robust Gesture Recognition	45
Warszawski,K.K., Nikiel,S.S.: Particle systems-based riverbed modelling over a terrain with hardness layer	49
Dundas,J., Wagner,M.: Adaptive Projection Displays: a low cost system for public interactivity	55
Giang,K., Funk,M.: Connect-S: A Physical Visualization Through Tangible Interaction	61
Kovačić,K., Ivanjko,E., Gold,H.: Real time vehicle detection and tracking on multiple lanes	67
Nowinski,K.S., Borucki,B.: VisNow - a Modular, Extensible Visual Analysis Platform	73
Simons,A., Prakash,E., Wood,J.: A new way of Rich Image Representation (VectorPixels)	77
Jang,H.-J., Baek,S.-H., Park,S.-Y.: Lane Marking Detection in Various Lighting Conditions using Robust Feature Extraction	83
Kim,J.-W., Kang,K.-K., Lee,J.-H.: Survey on Automated LEGO Assembly Construction	89
Bukatov,A.A., Gridchina,E.E., Zastavnoy,D.A.: A Control Cluster Approach to Non-linear Deformation	97

Ocean wave simulation by the mix of FFT and Perlin Noise

LI TIAN

Institute of Literature, Chinese Academy of Social Sciences
China, Beijing
skyhumxp@163.com

ABSTRACT

For the application of ocean wave, a new height-field simulation method is proposed by the mix of FFT and Perlin Noise, and OpenSceneGraph (OSG) and VC++ 2008 are used to simulate realistic ocean wave. The method takes wind effects into consideration, and uses Philips model and Gauss random numbers to produce ocean wave spectrum, which is then transformed to wave height-field by FFT. Perlin Noise is overlaid to disturb the wave height to generate a vivid and random sea surface. Simulation results show the effectiveness of the proposed method.

Keywords

Ocean wave spectrum, Fast Fourier Transform (FFT), Perlin Noise, Height-field, OpenSceneGraph (OSG)

1. INTRODUCTION

The ocean covers 71 percent of the Earth's surface, and plays a significant role in the Earth's environment. Ocean wave simulation has attracted a lot of attention since the appearance of computer graphics. It can be used in virtual reality, digital Earth, film art and so on [Dar11a]. For example, in Titanic and The Day After Tomorrow, many sea water scenes were simulated by computer [Tes01a].

Nowadays there have been several ocean wave simulation methods, which can be classified into five different types. The first type is based on physical model, the second is based on geometry model, the third is based on wave spectrum, the fourth is based on particle system, and the fifth is based on Perlin Noise.

The first type uses the classical fluid mechanics and solves Navier-Stokes equations to get realistic wave shape, but it is too time-consuming [Fos00a]. The second simulates wave shape by constructing a mathematical function, and its effect is highly dependent on the function itself [Fou86a] [Tso87a]. The third assumes that the ocean wave is composed by many sine waves with different amplitudes, different frequencies and different propagating directions, and FFT are used to transform wave spectrum to the corresponding ocean wave [Tes01a] [Mit05a]. The fourth supposes the water surface is made up of a lot of moving particles, which follow certain rules. Large number of particles will formulate realistic wave, but also cause heavily computation burden [Pea86a]. The fifth uses the

summation of coherent noises to construct sea surface height field, the trouble is that it can't account for wind effect [Per85a].

As statistical ocean wave spectrum models are adopted, FFT method can generate high realistic wave. But it supposes the whole sea surface is impacted by the same wind, hence the generated ocean wave has some similarity and repeatability, especially for a large area of sea. The paper proposes a new ocean wave simulation method, which introduces Perlin Noise into FFT method to improve vitality and reality of ocean wave. Also in Windows platform, an example of realistic ocean is simulated by OSG and VC++ 2008.

2. HEIGHT-FIELD GENERATION

FFT method

With the assumption of no surface tension and uniform seawater density, the solution of simplified Navier-Stokes equations could be expressed as follows

$$h(x, y, t) = \iint \bar{h}(k_x, k_y, t) \exp\{i(k_x x + k_y y)\} dk_x dk_y \quad (1)$$

$$\bar{h}(\vec{k}, t) = \bar{h}_0(\vec{k}) \exp\{-i\omega_0(\vec{k})t\} + \bar{h}_0^*(-\vec{k}) \exp\{i\omega_0(\vec{k})t\} \quad (2)$$

$$\omega_0(\vec{k}) = \sqrt{g|\vec{k}|} \quad (3)$$

Where $h(x, y, t)$ is the wave height at time t and position (x, y) , $\bar{h}(\vec{k}, t)$ is the spatial spectrum of $h(x, y, t)$, and $\vec{k} = (k_x, k_y)$ is the two-dimensional ocean wave vector.

For a sea surface with area $L_x \times L_y$ and grid number $N \times M$, the digitized (1) would be

$$h_{\text{fft}}(x, y, t) = \sum_{n=N/2}^{N/2-1} \sum_{m=M/2}^{M/2-1} \bar{h}(\vec{k}, t) \exp\{i(k_x x + k_y y)\} \quad (4)$$

And

$$\begin{aligned} k_x &= 2\pi n / L_x, n \in [-N/2, N/2), \\ k_y &= 2\pi m / L_y, m \in [-M/2, M/2) \end{aligned} \quad (5)$$

The equation (4) can be computed by 2D FFT to improve its efficiency. In (2), Phillips spectrum is often used to calculate ocean spatial spectrum, so

$$\bar{h}_0(\vec{k}) = \frac{1}{\sqrt{2}} (\xi_r + i\xi_i) \sqrt{P_h(\vec{k})} \quad (6)$$

$$\sqrt{P_h(\vec{k})} = A \frac{\exp\left(-\frac{1}{(|\vec{k}|L)^2}\right)}{|\vec{k}|^4} |\vec{k} \cdot \vec{\omega}|^2 \quad (7)$$

Where ξ_r and ξ_i are Gaussian number with mean 0 and variance 1, A is the wave amplitude, $\vec{\omega}$ is the wind vector, and $L = \frac{|\vec{\omega}|^2}{g}$ is the maximal wave length under $\vec{\omega}$.

In Philips model, the cosine factor $|\vec{k} \cdot \vec{\omega}|$ eliminates the energy vertical to wind direction, and makes the main direction of generated wave consistent with wind direction. The maximal wavelength L is determined by wind speed, the larger the wind speed is, and the higher the generated wave will be. When the wavelength is very short ($l \ll L$), the convergence of Philips model will be very band, and the following formula can be multiplied to improve it.

$$\exp\left(-|\vec{k}|^2 l^2\right) \quad (8)$$

Perlin Noise method

In 1985, Ken Perlin developed Perlin Noise function, which used fractal principle to generate coherent

noise [Per85a]. Perlin Noise is rather successful to simulate landscapes, clouds and many different sorts of textures.

The ocean wave simulation method by Perlin Noise could be shown as

$$h_{\text{perlin}}(x, y) = \sum_{i=0}^{P-1} a^i \cdot \text{noise}(2^i x, 2^i y) \quad (9)$$

Where P denotes the number of octaves to be added, 2^i and a^i are the frequency and amplitude of the octave i respectively, and $0 < a < 1$. The more octaves, the more details will be.

Fig. 1 gives a typical example of Perlin Noise. The left image is the basic noise image, the next three images are its octave 2, 4, and 8, and the right image is the coherent summation of all left four images. Fig. 1 shows that special texture can be generated by the summation of multiple fractal images with different frequencies and different amplitudes.

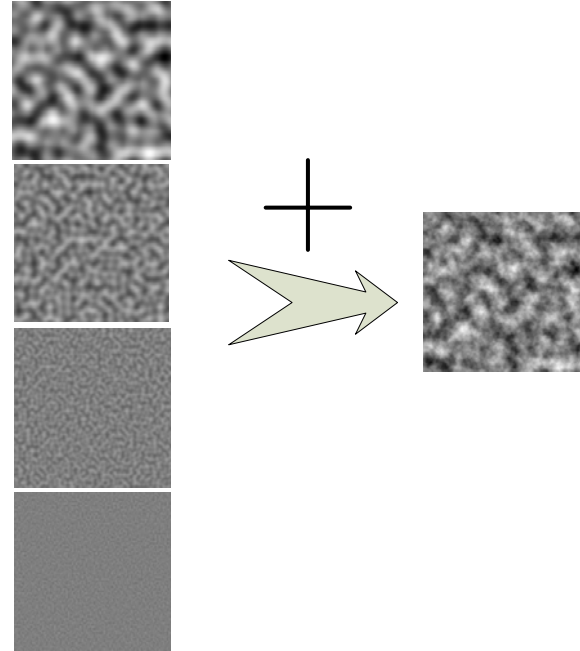


Figure 1 Perlin Noise

Mixed Perlin noise and FFT method

Wind is one of the most important reasons for ocean wave. As statistical and empirical spectrum models are adopted, FFT method can generate wave height-field with wind effect. But the behavior of wind is very complex, even in a small area, the wind speed and direction are not the same. When we see sea surface from near to far, clear wave direction by wind can be seen at near distance, but in the far distance, there is only noise like surface. Perlin Noise can also be used to generate sea surface, but it can't

consider the effect of wind. Therefore the fusion of FFT method and Perlin Noise will be an effective way to produce much more realistic wave.

The mixed height-field could be expressed as following

$$h_{\text{mixed}}(x, y, t) = \alpha(x, y) \sum_{n=N/2}^{N/2-1} \sum_{m=M/2}^{M/2-1} \bar{h}(\vec{k}, t) \cdot \exp\{i(k_x x + k_y y)\} + \beta(x, y) \cdot \sum_{i=0}^{P-1} a^i \cdot \text{noise}(2^i x, 2^i y) \quad (10)$$

Where $\alpha(x, y)$ and $\beta(x, y)$ are the weight coefficients of FFT method and Perlin Noise respectively. As the output of Perlin Noise is always between -1 and +1, its coefficient should be $\beta(x, y) \approx h_{0\max} [1 - \alpha(x, y)]$, where $h_{0\max}$ is the maximal value of $\bar{h}_0(\vec{k})$. By choosing different $\alpha(x, y)$ and $\beta(x, y)$, the proportion of FFT method and Perlin Noise can be controlled, therefore much more realistic ocean surface can be simulated.

3. OCEAN SURFACE SIMULATION

OpenSceneGraph (OSG) is an open-source 3D graphics engine. It provides many scene management and graphics rendering functionalities to the development of graphics programs, and has been wide-spreading applications in visual simulation, game, virtual reality, high technology research. OSG is programmed based the industrial standard OpenGL and ANSI C++, so it has the feature of cross-platform. By providing an object oriented framework using the concept of scene graph, OSG can free developers from the underlying invocation of OpenGL API.

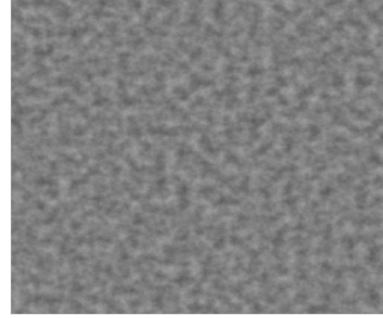
OSG library and VC++ 2008 are used to simulate ocean scene in this paper. Triangular-patch model of ocean surface is constructed based by the height-field data, and then realistic ocean surface can be simulated by color texture and its mapping. A simulation example will be given in the next section.

4. EXAMPLES

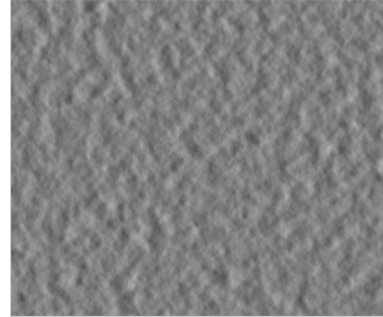
Fig. 2(a), (b) and (c) shows three ocean height-field images generated by Perlin Noise method, FFT method and the proposed method respectively. In these images, simulated ocean size is $512m \times 512m$, grid size is $1m \times 1m$, and pixel grayscale denotes the corresponding wave height. The brighter the pixel is, the higher the wave height is. For FFT method, wind speed and direction are 7m/s and 0 degree (from left to right) respectively. In order to show the wind effect in near field and noise

effect in far field, $\alpha(x, y) = 1 - \left(\frac{y}{L_y}\right)^3$ is used in

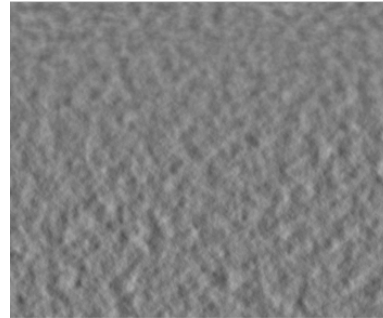
the proposed method to mix the FFT and Perlin Noise. Furthermore fig. 3(a), (b) and (c) depicts wind effect of 45° , 90° and 135° respectively. Fig. 4 is a rendering example of fig. 3(b) by OSG and VC++ 2008.



(a) Perlin Noise method

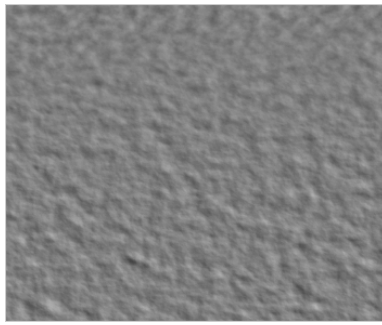


(b) FFT method

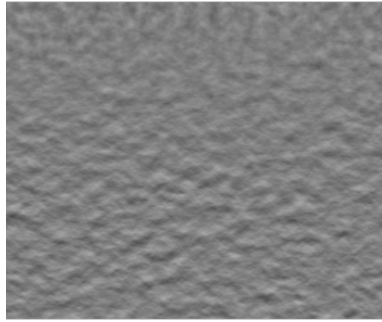


(c) mixed method

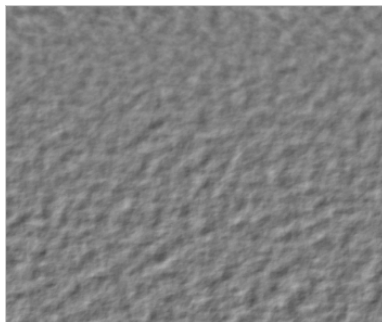
Figure 2 Sea height-field simulations by different methods



(a) 45°



(b) 90°



(c) 135°

Figure 3 Wind direction effect

5. CONCLUSION

As the development of computer graphics, ocean simulation has been acquiring more and more

application. The paper proposes a new ocean simulation method, which fuses FFT and Perlin Noise to generate a more vivid and random sea height-field. Simulation results are given to validate the proposed method; however the realistic rendering of large ocean scene and real-time accelerated processing will be the next study focus.

6. REFERENCES

- [Dar11a] Darles, E., Crespín B., Ghazanfarpour, D., et al. A survey of ocean simulation and rendering techniques in computer graphics. *Computer Graphics Forum* 30, 1, pp.43-60, 2011.
- [Tes01a] Tessendorf J. Simulating ocean water, *SIGGRAPH 2001: Inspire Interaction and Digital Images*, pp.301-318, 2001..
- [Fos00a] Foster N., Metaxas D. Modeling water for computer animation. *Commun ACM*, 43(7), pp.60-67, 2000.
- [Fou86a] Fournier A., Reevesw T. A simple model of ocean waves. *Computer Graphics*, 20(4), pp.75-84, 1986.
- [Tso87a] Ts O P., Barskey B. Modeling and rendering wave: wave-tracking using beta-splines and reflective texture mapping. *ACM Transaction on Graphics*, 16(3), pp.191-214, 1987.
- [Mit05a] Mitchell J. Real-time synthesis and rendering of ocean water. Marlborough: Array Technology Industry Technologies inc., 2005.
- [Pea86a] Peachy D. Modeling wave and surf. *Computer Graphics*, pp.65-74, 1986.
- [Per85a] Perlin K. An image synthesizer. *Computer Graphics*, 19 (3), pp.287-296, 1985.
- [Rui12a] Rui W., Xuelei Q. *OpenSceneGraph 3 Cookbook*. Birmingham: Packt Publishing, 2012.

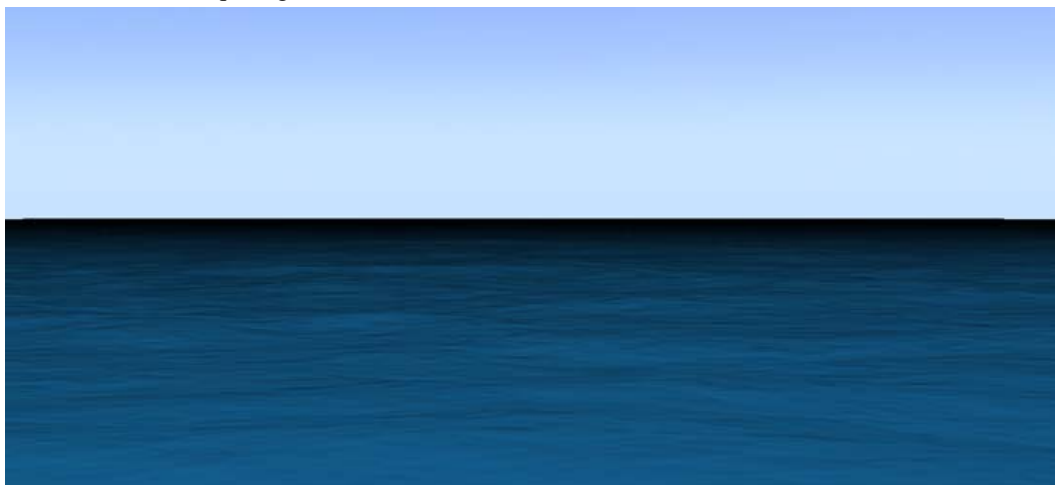


Figure 4 Simulation example

Visualizing multi-channel networks

Paweł Antemijczuk	Marta Magiera	Sune Lehmann	Andrea Cuttone	Jakob Eg Larsen
Technical University of Denmark DTU Compute Matematiktorvet 303B Denmark 2800, Kgs Lyngby, pantemijczuk@gmail.com	Technical University of Denmark DTU Compute Matematiktorvet 303B Denmark 2800, Kgs Lyngby, marmagiera@gmail.com	Technical University of Denmark DTU Compute Matematiktorvet 303B Denmark 2800, Kgs Lyngby, sljo@dtu.dk	Technical University of Denmark DTU Compute Matematiktorvet 303B Denmark 2800, Kgs Lyngby, jaeg@dtu.dk	Technical University of Denmark DTU Compute Matematiktorvet 303B Denmark 2800, Kgs Lyngby, ancu@dtu.dk

ABSTRACT

In this paper, we propose a visualization to illustrate social interactions, built from multiple distinct channels of communication. The visualization displays a summary of dense personal information in a compact graphical notation. The starting point is an abstract drawing of a spider's web. Below, we describe the meaning of each data dimension along with the background and motivation for their inclusion. Finally, we present feedback provided by the users (31 individuals) of the visualization.

Keywords

Multi-channel data, social networks, visualization, personal data, social interaction.

1. INTRODUCTION

Visualizing social connections is a recurring subject in the field of network science. Researchers like to view them in order to get a general overview of the network, before performing in-depth analysis, but even regular people like to view their own social networks, sometimes learning something about themselves, that they did not previously realize [1].

Visualizations most often take form of graphs with nodes denoting people involved in the network and edges showing the established social connections. This form is constantly refined and attempts are made to make it more clear and readable, especially for larger networks [2]. But simply visualizing nodes and connections between the users is not everything. Both people and their relationship, so both nodes and edges can have certain attributes. By drawing an edge, only one information is conveyed – that the two people know each other (unless it is a directed graph, in which one person may claim to know the other, but not vice versa). However a social connection is much more than a binary fact and by simplifying it as such, there is a loss of information

[3]. The connection can carry a wealth of data, such as channel or mode of communication (face-to-face, via a phone call, through a social networking site), the frequency of contacts as well as such things as geographical location of contact or the mood of the conversation. This information is potentially of high interest to both scientists and the participating subject.

In our approach to visualizing social connections, we would like to focus on smaller sub-networks, but attempt to visualize as many attributes of the connections as possible, while placing less emphasis on the node attributes. We hope that this will provide a unique insight on the user's own connections and provide interesting and stimulating self-feedback.

2. RELATED WORK

There are many tools that create visualizations of networks, such as have been mentioned before. Most of them focus on large, sprawling networks and attempt to display them in a clear way. ContactMap [4] is an example of one of the earliest attempts of sorting an individual's social contacts in a more concise, clear and organized way. However the majority of visualizations focus only on showing the structure of the network. While some tools attempt to encode various information using attributes such as color, positioning and shape of the drawn nodes [5][6] this information usually pertains to the structure of the network, such as the community they belong to [4], the amount of connections (the degree of the node) or network distance [5].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Alternatively, certain static information can be included, such as gender, organization the person belongs to, their city of residence, etc. Displaying all that information is difficult with the limited transformations that can be applied to a node, so other systems utilize panes to display different node attributes or draw expandable overlapping nodes for each of the node's attributes [7].

This still does not allow an individual to view attributes of the edges (connections), which Schneiderman et al. define as one of the six main challenges of network visualization [8].

Not many of these approaches focus on “multi-layered” social networks, (which is the case not only in visualization but also general network analysis) [9]. In such networks, different forms of contacts form different connections. The common approach is to construct several networks for each of the attributes and then combining them [3]. In principle this creates a network, which edges of have several attributes (or lack of thereof). The analytical approach does not make visualization any simpler and, in fact, combining several networks with use of color-coded edges to denote different types of links tends to be cluttered. Thus most tools fail to convey more attributes of both nodes and edges [7].

Solving the challenge, however, becomes easier as the size of the network decreases, especially if we are mostly interested in displaying detailed information about a single individual in the network (the subject of the visualization).

Additionally there is an increasing interest in self-quantified data and visualization that would aid in the process of learning about oneself [1] [10].

3. THE VISUALIZATION

In our visualization we focus on subsets (16 contacts) of an individual's social network, having built the said network out of complete information about the said individual – that is knowing about each and every social contact occurrence during a set period of time, its type (e.g. a phone call or a face-to-face meeting), duration, location and a detailed timestamp. We attempt to convey the most information possible about the individual's connections, narrowing them down to the most frequent ones or people that he or she had spent the most time with (as detailed later). Three data channels can be clearly seen in the dataset – contacts being made using the Bluetooth probe (implying face-to-face meeting), text messages as well as phone calls. Each of the three forms a separate social networks of contacts for the given user.

Instead of drawing a traditional graph we visualize the contacts in a different manner – as a metaphor of a spider's web. We would represent the user as the

“spider” sitting in the middle of the web and social contacts would be “caught in the web” in various locations of it, depending on certain parameters.

Figure 1 shows the final webchart. The chart displays only one channel.

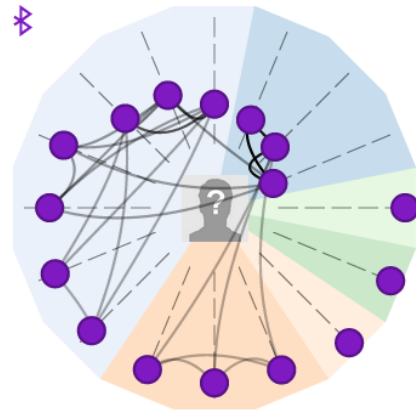


Figure 1. The final webchart displaying 16 friends of a user with which he made contacts using Bluetooth. The friends are linked based on their knowledge of each other and are grouped into communities.

As previously, each circle represents a single person the user has made contact with using the channel that the chart represents. In this case, they are Bluetooth contacts. In the middle of the chart there is a picture (or an avatar) of the user – for the testing purposes we used a placeholder picture displayed when the user does not have a picture in the system. The closer a circle is to the picture in the middle, the more contact has been made with the given person. The radial axes are meant to show the range of the values of the amount of contacts the user made with all his friends. The scale is between the minimal and maximal values for contact amounts between all friends within that channel.

The webchart is built with use of radial axes and links between the circles. The links represent the connections between the friends of the user themselves using the same channel. To draw these connections, first we construct a regular graph for all social contacts in the network excluding the user. After the graph has been created using all data, less significant links are removed using a thresholding algorithm described by Serrano et al. [11] This ensures we only show the most significant links.

After the network has been made and links pruned, community detection is ran on it, using our implementation of the Louvain method [12] based on the Python implementation in the NetworkX package¹. The background color of the axes is chosen based on the community the given friends belong to.

¹ <http://networkx.github.io/>

3.1.1 Small-multiple

The biggest challenge that we faced throughout the various iterations of the visualizations was how to display all three data channels for the user's social contacts. The data cannot be directly compared as each of the communication forms are different, however we would still like to show them simultaneously as well as allow people to make indirect comparisons between the contacts based on all three channels. In order to facilitate this we have decided to use the "small-multiple" concept as introduced by Tufte [13]. Instead of drawing one web, we would draw a web chart for each of channels used, as shown in Figure 2.



Figure 2. Small-multiple webcharts each for a different channel in the network – sms, Bluetooth contacts and phone calls.

This allows the user to compare their contacts across the channels, without introducing direct comparison between that data itself. This is possible as each contact remains on the same axis on each of the charts. In the figure we have concealed the community information for clarity. Note that the other channels do not have contact information at all; due to privacy reasons we are unable to show text messages and phone calls between the friends of the user. For Bluetooth, in fact, only contacts made in the presence of the user are recorded.

In case no data is not present for a channel, then the circle is not present on that chart.

3.1.2 Timeline

Below the multiple web charts we have placed a timeline that allows the user to change the period of time that the data is read for. If the period is changed, the circles slide across the axes to their new positions according to new data. Each chart shows the top 16 contacts (according to the sum of all contacts) for the given period.

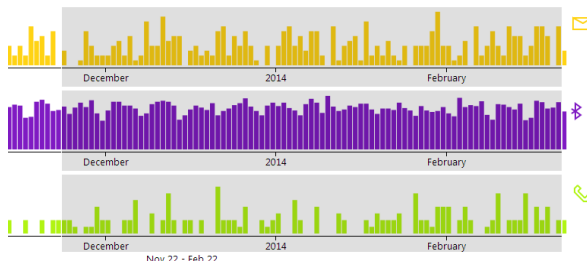


Figure 3. The timeline with barcharts for each channel showing total contacts made using the

given channel. The grey area signifies the time period chosen to be displayed on the webcharts.

The timeline is built out of bar charts – one for each channel. The bar charts show the total amount of contacts for the given; each bar being a single day. This provides a good overview of each channel's usage over time and allows us better to answer the main question about the user's data – "How does the user use his channels?"

4. EVALUATION

We have created test data sets in order to evaluate the visualization. The test data set that we tested against contained 16 participants grouped into 4 communities of friends that frequently meet during the weekday as well as 2 communities that represent friends meeting during the weekend. The communities slightly overlap. Additionally we have chosen a number of users to have much higher rate of contact using each channel.

After loading the data in the visualization the communities were detected correctly, as well as top contacts. We could clearly see that the test groups we have created in our data generators had the same background color on the visualization, that signified the communities. The contacts we have given the highest probabilities in the test generator, surfaced as closest to the center of the visualization.

4.1.1 User feedback

We have distributed the visualization along with an online survey to 31 people out of our friends and acquaintances. We asked them to evaluate the visualization, identify its features and the information that is conveyed.

In the first set we asked the users to rate on a scale from 1 to 5 (where 1 is the worst, 3 neutral and 5 the best) whether the visualization is: clear and understandable, easy to use, fun, novel and working as expected. The users responded mostly positively on all those questions with 74.2% finding it clear and understandable, 83.87% finding it easy to use, 58.06% fun, 80.64% novel and 77.42% working as expected. 9.68% of respondents found it completely unclear.

In the second section we asked the users to identify what information is conveyed by the visualization. We asked about several things, as well as some information that is **not** conveyed by the visualization in order to identify correct answers. The breakdown is as follows:

- 55% were able to identify their best friends.
- Almost 95% identified which of the contacts is called the most.
- 82% correctly identified in which days the user texts the most

- 63% correctly identified which contacts know each other
- 53% correctly identified which friends are good friends with each other (implying the community)
- 66% correctly identified what fraction of total contacts in given day, some chosen contacts are

Interestingly, even though no direct comparison are made, 79% of respondents identified which channel is used the most. This can be inferred from the popup information displaying the details, as well as the bar chart being far more uniform (contacts are made every day using that channel, not so much others). This allows for certain comparisons being made, without any direct comparisons.

There was a number of people who identified incorrect information, such as location of meetings, which implies they did not understand the visualization at all.

Lastly we asked them for any comments. In general the feedback was positive. Additionally other small suggestions were made regarding colors used and small clarifications.

5. CONCLUSIONS

We have successfully created a visualization of social contacts that is able to display three communication channels at once. It allows the user to display three (although there is nothing that would prevent from this model being used for more channels) different layers of their social network including detailed information about them.

The feedback was largely positive, with the only remarks being about the discovery of some features. This implies certain cosmetic changes might be necessary to make some features (especially the timeline changes) easier to discover.

This model can be successfully used to display multi-channel social networks, while our data contains only three channel and full interactions for only one of them, it is easily adapted for much more.

REFERENCES

- [1] I. Li, Y. Medynskiy, J. Froehlich and J. E. Larsen, "Personal informatics in practice: improving quality of life through data," *CHI '12 Extended Abstracts on Human Factors in Computing Systems, CHI EA '12*, 2799, pp. 2799-2802, 2012.
- [2] C. Dunne and B. Shneiderman, "Motif Simplification Improving network Visualization Readability with Fan and Parallel Glyphs," Human-Computer Interaction Lab Tech Report HCIL-2012-11, University of Maryland, 2012.
- [3] B. T. Dai, F. C. T. Chua and E. Lim, "Structural Analysis in Multi-Relational Social Networks," in *SIAM International Conference on Data Mining (SDM2012)*, 2012.
- [4] B. A. Nard, S. Whittaker, E. Isaacs, M. Creech, J. Johnson and J. Hainsworth, "Integrating Communication and Information Through ContactMap," *Communications of the ACM vol. 45 issue 4*, pp. 89-95, April 2002.
- [5] J. Heer and D. Boyd, "Vizster : Visualizing Online Social Networks".*Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization (INFOVIS '05). IEEE Computer Society, Washington, DC, USA, 5-* DOI=10.1109/INFOVIS.2005.39 <http://dx.doi.org/10.1109/INFOVIS.2005.39>.
- [6] U. Brandes and D. Wagner, "Analysis and Visualization of Social Networks," 2001.
- [7] A. Bezerianos, F. Chevalier, P. Dragicevic, N. Elmqvist and J. Fekete, "GraphDice: A System for Exploring Multivariate Social Networks," *Computer Graphics Forum*, 29 doi: 10.1111/j.1467-8659.2009.01687.x, p. 863-872, 12 August 2010.
- [8] B. Schneiderman and A. Aleks, "Network visualization by semantic substrates," *Visualization and Computer Graphics, IEEE Transactions on* 12.5, pp. 733-740, 2006.
- [9] P. Kazienko, P. Brodka, K. Musial and J. Gaworecki, "Multi-Layered Social Network Creation Based on Bibliographic Data," *Social Computing (SocialCom), 2010 IEEE Second International Conference* doi: 10.1109/SocialCom.2010.65, pp. 407,412, 20-22 August 2010.
- [10] J. Brophy-Warren, "The New Examined Life.," *The Wall Street Journal*, 6 December 2008.
- [11] M. Á. Serrano, M. Boguñab and A. Vespignanic, "Extracting the multiscale backbone of complex weighted networks," *PNAS vol. 106*, pp. 6483-6488, 21 April 2009.
- [12] P. D. Meo, E. Ferrarax, G. Fiumara and A. Provetti, "Generalized Louvain method for community detection in large networks," *CoRR, abs/1108.1502*, 2011.
- [13] E. R. Tufte, *The Visual Display of Quantitative Information*. Second edition, Cheshire, Connecticut: Graphics Press.

Parallel extraction hand vein biometric parameter's using a low cost IR imaging system

Sarah BENZIANE Abdelkader Benyettou Asmaa Merouane
 University of Oran USTO - MB University of Oran
 Algeria 31000, Oran, Es Sénia Algeria 31000, Oran, Bir el Djir Algeria 31000, Oran, Es Sénia
 Sarah_benziane1@yahoo.fr

ABSTRACT

This paper presents a new and low cost approach to authenticate individuals using hand vein images. The proposed method is fully automated and employs palm dorsal hand vein images acquired from a low cost , near infrared contactless imaging; the aim of our work. In order to evaluate the system performance, a prototype was designed and a dataset of 34 persons from different ages above 20 and different gender, each has 10 images per person was acquired at different intervals, 5 images for left hand and 5 images for right hand. The vein detection process consists of an easy to implement a device that takes a snapshot of the subject's veins under a source of infrared radiation at a specific wavelength. The system is able to detect veins but not arteries due to the specific absorption of infrared radiation in blood vessels. Almost any part of the body could be analyzed in order to extract an image of the vascular pattern but the hand and the fingers are preferred.

Keywords

Hand vein, Otsu, vascular biometrics, Infrared, dorsal vein, akka actors.

1. INTRODUCTION

The biometrics word has a larger meaning in the study of identification's persons from a number of characteristics[18]. A complex human inheritance, very rich in combinations, and perfectly adapted to such systems of user identification, and/or authentication [18][Ben11] [Wan05]. It's a Mathematical analysis of biological characteristics of a person to determine his identity decisively. Biometrics based on the principle of some characteristics recognition's. Fingerprints, face, iris, retina, hand, keystroke [Jai03] [Par05] and voice, provide irrefutable proof of the identity of a person they are unique biological characteristics distinguishing one person from another. The hand vein biometrics has emerged as a promising component of biometrics study.

Prior work

Such system was invented by the British engineer Joe Rice in 1984. Before the discovery to market, a study was commissioned at Cambridge Consultants to establish that the veins of the hand are unique to each individual.

The user places his hand in a room or a gauge reading. The characteristics of the veins are read by an infrared camera that takes a two-dimensional image [Ben12].

The hand vein biometrics principle is a noninvasive, computerized comparison of subcutaneous blood vessel structures (veins) in the back of a hand to verify the identity of individuals for biometric applications. Vein check measures are the shape and size of veins in the back of the hand (or front of the wrist). The vein pattern is best defined

when the skin on the back of the hand is taut – when the fist is clenched.

The captured images contain not only vein patterns but also irregular shading and noise. The shading is produced by the varying thickness of the muscles. Therefore, regions in which the veins are and are not sharply visible exist in a single image. It is widely known that the thermal (far IR) imaging cameras are highly sensitive to ambient conditions and very expensive. Therefore the researchers [Par05]-[Wan05], [Tan06], [Par97] [Hua12] have focused on the solutions using near IR imaging. Table 2 presents a comparative summary of prior work on the hand vein (back surface) approaches presented in the literature.

Table 1: comparative summary of related work, on hand vein

Reference	Methodology	Imaging	databases
T Tanaka & N.Kubo [Tano6]	FFT based phase correlation	Near Infra Red.	25 users
C-L.Lin & K.C.Fan [18]	Multi resolution analysis and combination	Thermal hand vein imaging (3.4-5um)	32 users
L.wang and G.Leedham [Wano5]	Line segment Hausdorff distance matching	Thermal hand vein imaging (8-14 um)	12 users
Y.Ding, D.Zhuang and K.Wang [9]	Distance between feature points	Near IR Imaging	48 users
J.M.Cross & C.L.Smith [Jai03]	Sequential Correlation in vein maps	Near IR Imaging	20 users
A.Kumar & K. Venkata	Matching vein and triangulation	Near IR Imaging	100 users

Prathyusha [Kumog]	shape features	contactless	
-----------------------	----------------	-------------	--

Proposed system

In this paper, we develop a new hand vein authentication approach using the structural similarity of hand vein shape features. The block diagram of the proposed approach is shown in figure

The efficacy of the proposed method is investigated from the trial results on the real hand vein images acquired with contactless near infrared imaging. Hence the steps of image normalization and feature extraction have been studied to decrease the variants in the interclass corresponding scores.

2. IMAGE ACQUISITION

The hardware configuration has a crucial role in the frame grabbing of veins. Two aspects can be underlined. The real camera used to take the stereotype has only one important parameter, the answer to the near infrared radiation [16]. The space resolution and the framework rate are of less importance since, for the acquisition of a model of vein, the image is necessary and the details are easily visible, even to a lower resolution. In [Cri] the design of the lighting system is one of the most important aspects of the process of frame grabbing. A good lighting system will provide precise data contrasts between the surrounding veins and fabrics while keeping Illuminations of the errors to a minimum. For a lighting system, they used IR LED because they offer a high contrast and easily available. But with a LED table formed using IR LED, they do not give a uniform illumination. Various arrangements were needed stamps LED to modify lighting. The LED is laid out like a simple or double table in 2D or rectangular table or concentric networks. Concentric arrangement table LED gives a better distribution of the light with only one or more concentric network of the LED and lens of the camera in the center of the image can acquire good contrasts.

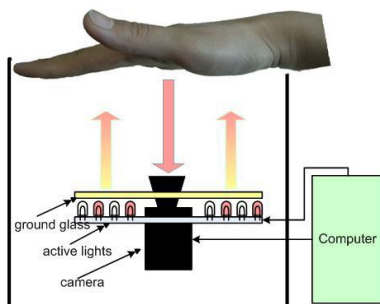


Figure 1 : Other system similar but which collects the veins of the fingers[HIT06].

3. SYSTEM DESIGN

For the realization of our hardware framework, we noted that the system needs some conditions in order

to function correctly such as: The day light influences on the quality of the image obtained except absence of IR filter; The temperature of the environment also influences on the quality of the image, it must be ambient neither too cold heat nor around the temperature of the human body; The distance between the sensor and the object must be sufficient for a good acquisition.

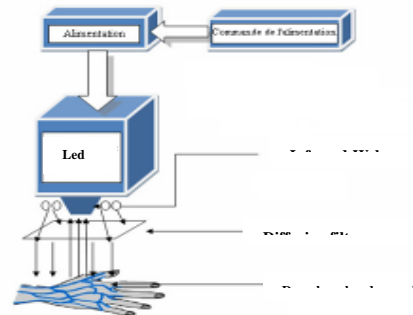


Figure 2 Hardware configuration

Our biometric acquisition system (SAB 11) is made with a case of wood; a modified webcam sensitive to the infra-red 850-900 nm waves; leds infra-red IR positioned in parallels. The transmission of information between the hardware and the software is done by an ordinary USB webcam cable.

The electric supply of the leds direct current (DC) is not a problem, it is not necessary to make an external supply circuit, it will be directly made through USB cable.

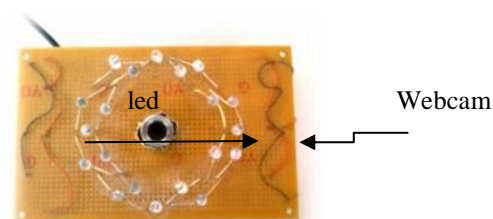


Figure 3 LEDS IR into concentric

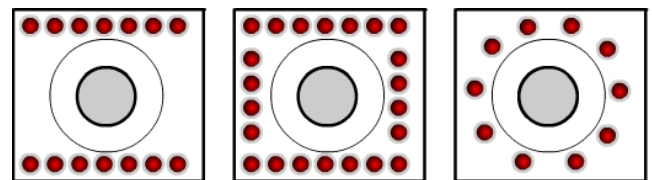


Figure 4 Various designs of the leds

Webcams are indeed sensitive at the same time visible spectrum of light as well as spectrum IR of the light. The filter outputs only the light visible with




camera. This filter is equipped either close to lens or on the case to the chip.

A special attention must be taken while withdrawing the filter IR on the case of chip because it can damage in a permanent way the webcam. Images captured to leave IR webcam significant are indicated below in the figure. We can see that by the IR LED illumination single used in remote television can provide illumination sufficient.

First open the webcam and to start by removing the anti-infra-red filter which does not let us see the IR. Take film of photography (Negative) and cut a circle with a diameter of 1 cm. Place the circle cut in the place of the removed filter and close the whole. Now, the webcam is ready for .

- After the modification of the webcam, and carrying out the test to collect IR images, we carried out a test on the hand Fig8.

- Our choice was to take an image of the back of the hand with a layout of the leds into concentric in circle with day light Fig 6.

		
Figure 5 Capture with webcam IR	Figure 6 capture with IR Webcam & IR LED	Figure 7 Capture in the case

For our first tests, we put the webcam at a distance of 17 cm of the hand and with only one circle of IR led of number of 10 surrounding the webcam, but we obtained an image not clear and with a black circle in the medium.

Number of image	Intensity current (A)	Tension (V)
1	0.39	1.5
2	0.3	1.4
3	0.17	1.3
4	0.1	1.2
5	0.06	1.2

Table 2: Various tests

Then, we choose a tension of 1.2 V with a current of 0.06 has with which one had a good catch. Then we carried out a test by doubling the circle with 20 led, but there was always the same problem of the black circle in the medium. Fig 9

Then, if we continue to add led that will never finish, one thought of making increase the distance between the webcam and the hand has 23cm, which gave us the following result. Fig10



Figure 8 Image 5



Figure 9 Image 6



Figure 10 Image 7






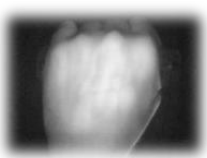
Figure 11 image 8

The conception finally finished after so many modifications and adjustments, the veins are visible with the IR imaging, then we can now start to make return a database of the people.

4. DATABASE

After systems design, we carry out the acquisition of the images of various people at end to have a database which will be used thereafter for the identification of the people; which are all evaluated on a common database with a common protocol.

To create the database we have first to connect SAB11 to the PC and launch the software. After, we open the door, set the hand on the wrist and close the door. Click on the button to get the image transferred to our framework. To make the capture thus five times for the right hand for the left hand. With taking

	Right hand	Left hand
First person		
Second person		

images so much the database will be effective.

Figure 12 : Sample of the database

If we compare each hand of only one individual we clearly see with the eye naked that they differ on the level from veins. If we compare two hands of two people with the eye naked we see that they also differ as for only one individual

But the identification is not also easy it is necessary to treat the images, i.e. to follow several steps extraction of the object, extraction of the veins so on.

Some step process (Laplacian of Gaussian) can be quite effective for the localization of venous structure in the acquired hand vein images.

The database has match scores corresponding to three hundred and forty files, which were acquired from trent-four people during one week. Mainly students (subjects between 20 and 30 years old) within the Institute of Maintenance and Industrial safety (University of oran Es Sénia). In each session, five recordings were taken.

5. FEATURE EXTRACTION FROM DORSAL HAND VEIN

In this part, we present a binarization algorithm for automatic extraction of the bottom region from the image adapted to a set of dorsal veins biometric images of the hand; in order to evaluate identification / authentication. We focus in this study on optimizing the time binarization images gray level, because what essentially makes a biometric system is efficient and effective is processing reply time.

We have used OTSU's method [18] as a global thresholding because the images of dorsals hand veins have equal illumination then global methods can work better. We worked on database composed of multiple images; right and left hand veins (10 to 20 images). In our case, we work with databases of 20 images per person 640 * 480 pixels. In order, to perform our biometric system for binarization step for several images, we proposed a parallel implementation of OTSU's algorithm to speed up the processing.

6. OTSU BINARIZATION'S METHOD

The aim of Otsu's method [16] is to determine the optimal threshold T from the image histogram. Therefore, the calculation of image histogram is made in first time, by counting for each gray level, the number of pixels associated.

Then, to make the process independent of the number of points in image, the histogram must be normalized as follow [Ben12] :

$$Normalize_histogram(i) = \frac{n_i}{255} \quad (1)$$

n_i represents the number of pixels of graylevel i of input image.

The separation into two classes (background and hand) is done by the mean and variance.

After that, for each value of k ($k \in [1, 255]$), we compute s :

$$s^2(k) = var(k) * (1 - var(k)) * (mean(255) * var(k) - mean(k))^2 \quad (4)$$

$$(2)$$

Finally the threshold value T is obtained when, for a given k ($k \in [1, 255]$) we have:

$$s^2(k) = \max(s(k)^2) \quad (555) \quad (3)$$

Now, we threshold image:

The pseudo-code of OTSU's method is as follow:

Compute histogram by calculating the number of pixels for each graylevel

for ($i=0$ to 255) {

Normalize_histogram(i)

}

Initialize $mean(0)=0, var(0)=0$

for ($k= 1$ to 255) {

update *mean(k)*

update *var(k)*

compute *s²(k)*

}

Threshold value is k where $s^2(k) = \max(s(k)^2) \quad (7)$

After the sequential implementation of OTSU's method for thresholding of 20 images of definition (640*480), we noticed that the execution time is full grown (6seconds) .Since in biometric ,we identify people in real time, we have optimized this step by using master-worker paradigm where thresholding's process still be divided over each processes (workers) in parallel. optimized this step by using master-worker paradigm where thresholding's process still be divided over each processes (workers) in parallel.

7. PARALLEL OTSU'S BINARIZATION

In order to binarize several images in parallel with OTSU's method we have opted to master/worker paradigm, where the master P0 creates several workers and dispatch for each one the number of image. After, each worker can subdivide the image into several blocks (sub images) and dispatch each block to other workers in order to threshold these sub images. The figure bellow shows our approach.

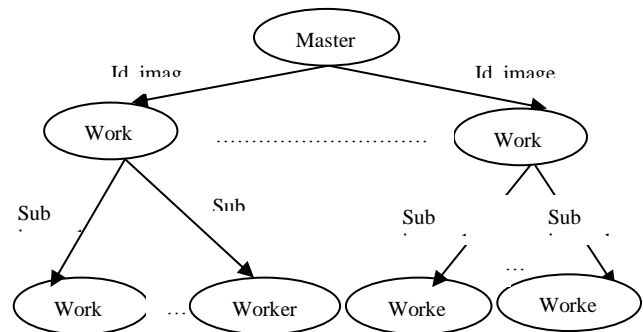


Figure 13. Master worker paradigm of binarization of several images (scatter phase)

In scatter phase, each worker of level 2 takes a sub image and applies OTSU's binarization method.

Person	Left/Right hand	Input image	Output image of OTSU's binarization
1	Left		
	Right		
2	Left		
	Right		

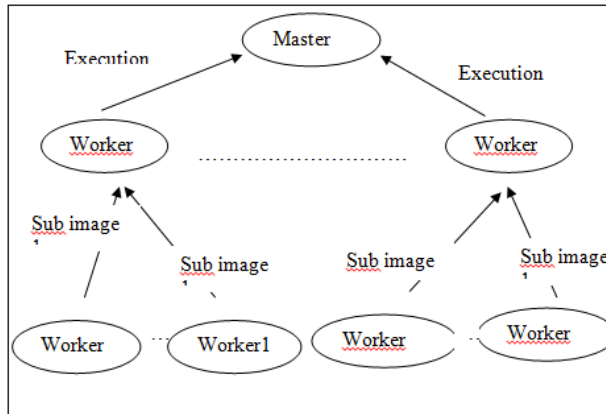


Figure 14. Master worker paradigm of binarization of several images (gather phase)

In gather phase each worker return its results resultants to its master, that will assembles the final results.

8. ROI EXTRACTION AND ENHANCED

To extract the area that contains only information of the vein pattern, we have calculated the center of

gravity (x_g, y_g) by (1) as made in [Kum09]. As the vein images were not clearly distinguished, we improved the contrast by linear transformation.

$$x_g = \frac{\sum_{i,j} i \times f(i,j)}{\sum_{i,j} f(i,j)} ; y_g = \frac{\sum_{i,j} j \times f(i,j)}{\sum_{i,j} f(i,j)}$$



(a). ROI of the vein images before enhancement



(b). ROI of the vein images after enhancement

9. HAND VEIN SEGMENTATION

Global thresholding method is not a good technique for this purpose [Cri07]. However the local thresholding is better approach for hand vein segmentation. In this study, in order to minimize the computational time of local thresholding calculation, we have used integral image [Cro95] [13] as local thresholding technique to separate the vein from background.

The main idea of the proposed method is to compute the integral image of each pixel [16] $g(x,y)$ by (2).

$$g(x,y) = \begin{cases} I(0,y) + g(0,y-1) & \text{if } x=0 \text{ and } y \in [1,m] \\ I(x,0) + g(x-1,0) & \text{if } y=0 \text{ and } x \in [1,n] \\ I(x,y) + g(x,y-1) + g(x-1,y) - g(x-1,y-1) & \text{else} \end{cases} \quad (1)$$

Where $I(x,y)$ is the intensity value of pixel (x,y) .

Once we have computed integral image, the local sum $s(x,y)$ [16] can be defined as follow.

$$s(x,y) = g(x+d,y+d) + (x-d-1,y-d-1) - [g(x-d,y-d-1) + g(x-d-1,y-d)]$$

10. HARDWARE AND SOFTWARE SPECIFICATION

Experimentations have been carried out on a laptop with an Intel Core 2 Duo/2GHz CPU and 3GB of RAM. Implementation has been done with Scala language version 2.9 and Akka actors version 2.0.5.

Scala is a functional and Object-oriented programming. It has parallel collection libraries for parallel processing of independent data. Since, Akka is Open Source and available under the Apache 2 License, it provides API for Java and Scala. We have used it because it offers a set of tools to solve

problems [Fuk] related to competition: actors, agents, software transactional memory (STM) and proposes mechanisms for fault tolerance. Akka Actors are very lightweight concurrent entities. They process messages asynchronously using an event-driven receive loop.

So, for our approach of master/worker paradigm, the master and all workers become actors that will communicate by message.

11. IMPLEMENTATION AND EXPERIMENTATION

We have tested our approach for the thresholding of 20 images in parallel with dimensions 640*480. The comparison of execution time of serial implementation over parallel is shown in table 1.

We have used one actor as a master and 20 Akka actors as workers of level one. In level two, each actor of level one can create several actors to binarize each sub image.

But the problem was in the choice of the number of actors to binarize an image. So the table bellow presents the results of experiments that we have run.

12. CONCLUSION

This paper has proposed an effective approach for the image acquisition of the dorsal vein patterns; according to experiments done for comparing our methods with other results. Filtered image quality is evaluated in order to find the best infrared wavelength for palm vein pattern imaging. After calculating the effectiveness index in images, it was shown that the best image quality is with the reflection method when using 760nm IR LEDs with optical filter.

Experiments with transmission method gave better results than the reflection method. However, with 760nm filter results were almost same as with the reflection method.

In this research work, a new approach based on actors for parallel implementation of OTSU's binarization for 20 images of hand vein, based on communication message between actors under the environment Scala and actors Akka, allowed us to obtain satisfactory results, however, the performance of this system can be still limited because of the communication time between actors becomes disadvantageous when the number of processes is important. So, we can say that Akka actors and Scala can be used for optimization in biometric or other situations of compute intensive or fault tolerance because Akka provides mechanisms for fault tolerance.

As a future work, we propose to use Akka actors on GPU, may it gives us more performance then the cpu for OTSU's binarization in dorsal hand vein

biometric. We can suggest too, implementing this step in FPGA and comparing the performance with Akka actors in GPU.

13. REFERENCES

- [Ben11] S.BENZIANE, A.BENYETTOU, « An Introduction to biometrics », International Journal of Computer Science & Information Security, IJCSIS Vol. 9 No. 4, April 2011 ISSN 1947-5500
- [Wan05] L. Wang and G. Leedham, "A thermal hand-vein pattern verification system," Pattern Recognition and Image Analysis, S. Singh, M. Singh, C. Apte and P. Perner (Eds), LNCS, vol. 3687, pp. 58-65, Springer, 2005.
- [Jai07] A. K. Jain, Y. Chen, and M. Demirkus, "Pores and Ridges: Fingerprint Matching Using Level 3 Features," IEEE. Tran. Patt. Anal. Mach. Intell., vol. 29, no. 1, pp. 15-27, 2007.
- [Par05] K. R. Park and J. Kim, "A real-time focusing algorithm for iris recognition camera," IEEE Trans. Sys. Man & Cybern., Part C, vol. 35, pp. 441-444, Aug. 2005.
- [Tan06] T. Tanaka and N. Kubo, "Biometric authentication by hand vein patterns," Proc. SICE Annual Conference, Yokohama, Japan, pp. 249-253, Aug. 2004.
- [Par97] G. T. Park, S. K. Im, and H. S. Choi, "A person identification algorithm utilizing hand vein pattern," Proc. Korea Signal Processing Conference, vol. 10, no. 1, pp. 1107-1110, 1997.
- [Tan04] T. Tanaka and N. Kubo, "Biometric authentication by hand vein patterns," Proc. SICE Annual Conference, Yokohama, Japan, pp. 249-253, Aug. 2004.
- [Din05] Y. Ding, D. Zhuang and K. Wang, "A study of hand vein recognition method," Proc. IEEE Intl. Conf. Mechatronics & Automation, Niagara Falls, Canada, pp. 2106 - 2110, Jul. 2005.
- [Kum09] Ajay Kumar, K. Venkata Prathyusha, "Personal Authentication using Hand Vein Triangulation and Knuckle Shape", IEEE Transactions on Image Processing, September, 2009
- [Cri07] S. Crişan, I.G. Târnovan, T.E.Crişan, „Vein pattern recognition. Image enhancement and feature extraction algorithms”, 15th IMEKO TC4 Symposium, Iaşi, Romania, 2007, ISBN 978-973-667-260-6 15
- [Paq06] Vincent Paquit, Jeffery R. Price, Near-infrared Imaging and Structured Light Ranging for Automatic Catheter Insertion, Medical Imaging OAK RIDGE NATIONAL LABORATORY, Oak Ridge, Tennessee, 2006
- [Fuk] "Infrared Imaging System for Analysis of Blood Vessel Structure", R. Fuksis, M. Greitans, O. Nikisins, M. Pudzs, Institute of Electronics and Computer Science.
- [Chi04] Chih-Lung Lin and Kuo-Chin Fan, "Biometric Verification Using Thermal Images of Palm-Dorsa Vein Patterns" IEEE Trans. Circuits & Sys. For Video Technology, vol. 14, pp. 199 - 213, Feb. 2004.
- [Ben12] S.BENZIANE, A.BENYETTOU, "State Of Art: Hand Biometric", International Journal of Advances in Engineering & Technology, IJAET, Vol. 2, Issue. 1, Jan-2012, ISSN: 2231-1

METHOD OF DISCRETE WAVELET ANALYSIS OF EDGES ON THE RANDOM BACKGROUND

D.A. Bezuglov

Don state technical university

Russian Federation
Gagarina Sq. 1
344016 Rostov-on-Don
bezuglovda@mail.ru

Yu.D. Bezuglov

Don state technical university

Russian Federation
Gagarina Sq. 1
344016 Rostov-on-Don
ybezuglov@gmail.com

S.A. Shvidchenko

Don state technical university

Russian Federation
Gagarina Sq. 1
344016 Rostov-on-Don
sveta_more@mail.ru

ABSTRACT

Currently known algorithms suppose preliminary image filtering with consecutive solving of edges extraction problem. In the process of development of image filtering algorithms the a priori knowledge of distortion interference characteristics is also required. In practice in most of the cases such information is not available or is approximate. A new method of objects' edges extraction in the images was developed, in the presence of distortion, by applying direct and inverse wavelet transformation. The results of mathematical modeling are provided. The proposed approach may be used in the course of development of digital picture processing systems, self-contained robots engineering, and under the surveillance conditions which complicate the registration process in the absence of a priori data about the background noise type.

Keywords

Edge detection, image analysis, wavelet-transformation.

1. INTRODUCTION

In recent times the systems of different objects identification are actively developed. Thereat the volume of stored information and its fidelity show notable growth. At the same time the problem arises of in-line data processing and valid data retrieval from large scale arrays of images. Such problems arise in quite a few brunches of knowledge: healthcare, radiolocation, space studies and Earth exploration, television, etc. For example, diagnostics by analysis of human innards image, forest fire detection, navigation objectives, synthetic vision for special systems, etc.

One of development trends of modern IT-technologies is the development of methods and algorithms of analysis of signals (measurement results) and their derivatives against the registration noises background. Such problems arise at mathematical simulation of different dynamic

processes and objects, described by differential equations in the course of control automation of given processes, and in the course of image processing in edge extraction problem.

Without the skill and knowledge of efficient solution of such type of problems, there is no sense to discuss elaboration of the corresponding systems of signals and images processing [Gon01a, Kra01a]. One common point of all the above problems is that a calculation of derivatives of different orders (as a rule, the first and the second) against the noise background is required for the automated processing of measurement results. The problem of signal differentiation is incorrect, that is why basing on traditional analogous differentiating circuits and amplifiers, it turns out that it is impossible to create an ideal differentiator or a differentiator, which is sufficiently close to it.

Solution of the edge extraction problem is used in industrial applications for engineering of self-contained robots, as well as the systems of image analysis under adverse surveillance conditions, various hindering factors, complicating the process of image registration in the absence of a priori data about the background noise type. It means that the methods and algorithms of data processing from the image detectors shall take into account the presence of noises of different nature, related to images and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

signals registration in actual systems. The above makes it absolutely apparent the significance of analysis of existing and development of new methods of digital differentiating of signals and images, registered at the noise background, and selection of such a single or several of them, which are best suited for implementation with the use of modern microprocessor engineering and providing for the needed characteristics, and not requiring the information about a priori characteristics of interference and background noises.

Hence, today the scientific task of development of the automated analysis methods for the results of measurements for objects' edge extraction in images in the presence of background noise currently is not properly solved, and is important.

2. PROBLEM SOLUTION

Let us examine in greater detail the algorithms of wavelet differentiation with the use of MHAT, DOG, WAVE wavelets, obtained on the basis of the earlier developed method of wavelet differentiation [Bez01a - Bez03a].

In order to find the edge it is necessary to calculate the gradient [Gon01a, Kra01a].

In general terms the derivatives of a row and a column of images may be expressed as follows:

$$\frac{\partial S(j)_i}{\partial x} = S1(j)_i = C_\psi \sum_{m=1}^{NK} \sum_{n=0}^{N-1} CTWSS(m, n, i) \frac{\partial \varphi(x_j)}{x}, \quad (1)$$

$$\frac{\partial S(i)_j}{\partial y} = S1(i)_j = C_\psi \sum_{m=1}^{NK} \sum_{n=0}^{N-1} CTWSC(m, n, j) \frac{d\varphi(y_i)}{dy}, \quad (2)$$

$$i = 0 \dots N, \quad j = 0 \dots N,$$

where $CTWSS(m, n, i)$ and $CTWSC(m, n, j)$ are correspondingly the coefficients of direct discrete wavelet transformation by the rows and columns of image matrix $S(i, j)$.

$$CTWSS(m, n, i) = \sum_{j=0}^{N-1} \varphi_{m,n}(x_j) S(i, j), \quad (3)$$

$$CTWSC(m, n, j) = \sum_{i=0}^{N-1} \varphi_{m,n}(y_i) S(i, j). \quad (4)$$

Then the expression for the gradient square of the matrix $S(i, j)$ shall appear as follows:

$$\begin{aligned} \left(\frac{\partial S(j)_i}{\partial x} \right)^2 + \left(\frac{\partial S(i)_j}{\partial y} \right)^2 &= \left(C_\psi \sum_{m=1}^{NK} \sum_{n=0}^{N-1} CTWSS(m, n, i) \frac{\partial \varphi_{m,n}(x_j)}{\partial x} \right)^2 + \\ &+ \left(C_\psi \sum_{m=1}^{NK} \sum_{n=0}^{N-1} CTWSC(m, n, j) \frac{\partial \varphi_{m,n}(y_i)}{\partial y} \right)^2 = \\ &= (S1(j)_i)^2 + (S1(i)_j)^2, \quad i = 0 \dots N, \quad j = 0 \dots N. \end{aligned} \quad (5)$$

The absolute value of the intensity gradient of image $S(i, j)$ under study in terms of wavelet transformation may be expressed as follows:

$$\begin{aligned} G(S(i, j)) &= \left(\left(C_\psi \sum_{m=1}^{NK} \sum_{n=0}^{N-1} CTWSS(m, n, i) \frac{\partial \varphi_{m,n}(x_j)}{\partial x} \right)^2 + \right. \\ &\left. + \left(C_\psi \sum_{m=1}^{NK} \sum_{n=0}^{N-1} CTWSC(m, n, j) \frac{\partial \varphi_{m,n}(y_i)}{\partial y} \right)^2 \right)^{1/2}. \end{aligned} \quad (6)$$

Novelty of the proposed approach is that in the expression (6) the wavelet functions derivatives of type $\frac{\partial \varphi_{m,n}(x_j)}{\partial x}$ and $\frac{\partial \varphi_{m,n}(y_i)}{\partial y}$ are used in reverse

wavelet transformation.

Images in the process of their forming by imaging systems (photographic, holographic, television) are usually affected by different random interferences or noises. The most common type of interferences is random additive noise, statistically independent from signal. The additive noise model well describes the effect of film-grain noise, quantization distortion in analog-to-digital converters, etc. That is the reason why in the course of mathematical simulation we shall use the random numbers generator, whereat the quality evaluation criteria of image edge extraction shall be statistical.

This paper uses the following criteria.

1. Root-mean-square deviation, e_{rmsd} :

$$e_{rmsd} = \sqrt{\frac{1}{MN} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \left(MT_{ij} - \hat{MK}_{ij} \right)^2} \quad (7)$$

Whereat the image of edges, obtained from a noise-free tested image S by the Canny edge detector, was used as the MT test image. Further on the additive Gaussian noise was superimposed on the source image S , and the edge extraction by the proposed wavelet differentiation method and by the known Sobel method was performed. At that the \hat{MK} images

were obtained.

2. Peak signal-to-noise ratio, $SNRI$:

$$\mu = \frac{1}{MN} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \left(\hat{MK}_{ij} \right)^2; \quad (8)$$

$$SNRI = \frac{255 - \mu}{\sqrt{\frac{1}{MN} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \left(\hat{MK}_{ij} - \mu \right)^2}}$$

where μ - mean value of \hat{MK}_{ij} .

3. Peak signal-to-noise ratio, $SNR2$ (with the use of background root-mean-square deviation (RMSD) in calculations):

$$\sigma_{bcgr} = \sqrt{\frac{1}{(N_{bcgr})^2} \sum_{i=n_1}^{n_1+N_{bcgr}} \sum_{j=m_1}^{m_1+N_{bcgr}} \left(M \hat{K}_{ij} - \mu_{bcgr} \right)^2}; \quad (9)$$

where: σ_{bcgr} - background RMSD;

$\mu_{bcgr} = \frac{1}{(N_{bcgr})^2} \sum_{i=n_1}^{n_1+N_{bcgr}} \sum_{j=m_1}^{m_1+N_{bcgr}} \left(M \hat{K}_{ij} \right)$ - mean value of the background;

n_1, m_1 – coordinates of the singled out background

area with dimensions $N_{bcgr} \times N_{bcgr}$ on the image \hat{MK} under study,

$$SNR2 = \frac{255 - \mu}{\sigma_{bcgr}}. \quad (10)$$

Using the three above criteria it shall be further possible to evaluate more adequately the efficiency of the proposed algorithms as compared to those already known.

For qualitative and quantitative evaluation of the developed algorithms let us test their functioning on test images. Whereat we shall study the test images, which were not affected by noise. See the Pictures 2÷7 for the results of mathematical simulation.

The procedure of computing experiment was as follows. The source eight-bit image with dimensions 512*512 “Cameraman” (Fig. 1), was processed with canny edge detector from MathCAD (Fig. 2). Then the “Cameraman” image was treated with the additive Gaussian noise with RMSD $\sigma=5, 20, 30, 40, 50$.

See Fig. 3 for the image with RMSD at $\sigma=30$. For the results of edge extraction by the known Sobel method see Fig. 4. See Fig. 5÷7 for the research of the proposed algorithms of edge extraction by the use of DOG, WAVE and MATH wavelets at noise background for the models with the mathematical noise expectation $m_0=0$. See Tables 1-3 for the values of e_{rmsd} , SNR1 and SNR2 for the series of the studied test images.



Figure 1. Initial image (8-bit).

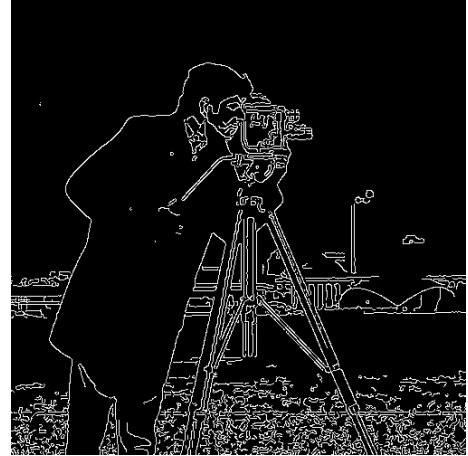


Figure 2. Initial image is processed by Canny edge detector.



Figure 3. Initial image with standard deviation $\sigma=30$.



Figure 4. Initial image is processed by Sobel operator.



Figure 5. Initial image is processed by DOG wavelet.



Figure 6. Initial image is processed by WAVE wavelet.



Figure 7. Initial image is processed by MATH wavelet.

RMSD of the forming noise	RMSD of source image, e_{rmsd}	Algorithms/ Criterion gain, dB		
		Wavelet DOG	Wavelet WAVE	Wavelet MATH
5,00	10,33	3,14	2,81	2,99
20,00	31,37	3,49	3,16	3,21
30,00	39,98	3,70	3,39	3,37
40,00	46,22	4,22	3,96	4,00
50,00	50,77	3,96	3,70	3,57

Table 1. Computing experiments results for criterion "Root-mean-square deviation"

RMSD of the forming noise	Peak signal-to-noise ratio, SNR1 of source image	Algorithms/ Criterion gain, dB		
		Wavelet DOG	Wavelet WAVE	Wavelet MATH
5,00	1,92	5,16	5,20	5,34
20,00	2,63	4,83	4,80	4,91
30,00	3,05	4,89	4,80	4,93
40,00	3,37	5,01	4,89	5,00
50,00	3,63	5,11	4,97	5,06

Table 2. Computing experiments results for criterion "Peak signal-to-noise ratio SNR1"

RMSD of the forming noise	Peak signal-to-noise ratio by RMSD of source image, SNR2	Algorithms/ Criterion gain, dB		
		Wavelet DOG	Wavelet WAVE	Wavelet MATH
5,00	8,11	9,64	8,86	9,59
20,00	5,94	7,51	7,63	7,16
30,00	5,40	7,10	7,32	6,66
40,00	5,07	6,91	7,20	6,43
50,00	4,89	6,81	7,14	6,30

Table 3. Computing experiments results for criterion "Peak signal-to-noise ratio SNR2"

The obtained quantitative results and expert evidence of image edge extraction allow for the conclusion about the advantages of the proposed algorithms as compared to already known. Software implementation of the proposed methods and algorithms allow for the automation of signals and images processing processes, and expand the possibilities of research conduct for development of advanced signals and images processing systems.

3. CONCLUSIONS

Analysis of the images and results of mathematical simulation shown on the figures allows us to make the following conclusions. The proposed method of

image processing allows to efficiently perform edge extraction of noise distorted images. The new method and the algorithms of wavelet differentiation of images at noise background with the use of discrete wavelet transformation allow to increase the peak signal-to-noise ratio by $4,8 \div 9,6$ dB and in $3 \div 4$ times decrease the error in standard deviation.

In the given case the wavelet transformation allow to refuse from the use of different masks, i.e. to refuse from inefficient methods of numerical differentiation. Using the proposed method the other algorithms of edge extraction on the basis of wavelet differentiation with the use of different wavelet bases may be implemented.

Hence, the scientific mission to elaborate the methods of automated analysis of measurement results for edge extraction of fragments of objects in images with the presence of a background noise is solved. The proposed methods and algorithms may be used at development of the signal processing systems of digital images for industrial application for self-contained robots engineering, for surveillance under the conditions hindering the registration process, and in the absence of a priori information about the type of background noises.

4. REFERENCES

- [Bez01a] Bezuglov D.A., Shvidchenko S.A. Information technology of wavelet differentiation measurements results on the background noise (article). // Vestnik komp'yuternykh i informatsionnykh tekhnologii (Herald of computer and information technologies). №6 (84) 2011. p.42-45.
- [Bez02a] Bezuglov D.A., Ritikov S.U., Shvidchenko S.A. Method of wavelet differentiation in the problem of allocation of contours (article). //Uspekhi sovremennoi radioelektroniki» (Achievements of Modern Radioelectronics). №6 2012r. p.52-57.
- [Bez03a] Bezuglov D.A., Ritikov S.U., Shvidchenko S.A., Gavrin C., Gavrin DS. Isolation of the contours image in information system and control system by using method wavelet transform (article). // Nelineinyi mir" (Nonlinear World) . № 11 2012r., p.846-852.
- [Gon01a] Authorized translation from the English language edition, entitled Digital image processing, 2nd edition by Gonzalez, Rafael C.; Woods, Richard E., published by Pearson Education, Inc, publishing as Prentice Hall, 2002.
- [Kra01a] Digital signal and image processing in radio physical applications. /V.F. Kravchenko. – M.: FIZMATLIT, 2007, - 544 p. – ISBN 978-5-9221-0871-3.

Polynomiography with non-standard iterations

Krzysztof Gdawiec
Institute of Computer Science
University of Silesia
Bedzinska 39
41-200, Sosnowiec, Poland
kgdawiec@ux2.math.us.edu.pl

Wiesław Kotarski
Institute of Computer Science
University of Silesia
Bedzinska 39
41-200, Sosnowiec, Poland
kotarski@ux2.math.us.edu.pl

Agnieszka Lisowska
Institute of Computer Science
University of Silesia
Bedzinska 39
41-200, Sosnowiec, Poland
alisow@ux2.math.us.edu.pl

ABSTRACT

In the paper the visualizations of some modifications applied to the Newton's root finding of complex polynomials are presented. Namely, instead of the standard Picard iteration, several different iterative processes described in the literature, which we call as non-standard ones, were used. Following Kalantari, such visualizations are called polynomiographs. Polynomiographs are interesting from scientific, educational and artistic points of view. By the use of different iterations kinds we obtain quite new polynomiographs, in comparison to the standard Picard iteration, which look aesthetically pleasing. We present some polynomiographs for complex polynomial equation $z^3 - 1 = 0$ as examples. Polynomiographs were defined to graphically present dynamical behaviour of different iterative processes. But we are not interested in that. We are focused on polynomiographs from the artistic point of view. We believe that the new polynomiographs can be interesting as a source of aesthetic patterns created automatically. They can also be used to increase functionality of the existing polynomiography software.

Keywords

polynomiography, iteration process, Newton method, computer art

1 INTRODUCTION

One can meet polynomials in many mathematical fields. They are interesting both from theoretical and practical points of view. Especially, the problem of polynomials root finding has a long and fascinating history. Already Sumerians 3000 years B.C. and ancient Greeks faced with practical problems that in modern mathematical language can be considered as a root finding of polynomials. In 17th century Newton proposed a method for calculating approximately roots of polynomials. Cayley in 1879 observed strange and unpredictable chaotic behaviour of the roots approximation process while applying the Newton's method to the equation $z^3 - 1 = 0$ in the complex plane. The solution of the Cayley's problem was found in 1919 by Julia. Julia sets became an inspiration for the great discoveries in 1970s – the Mandelbrot set and fractals [Man83]. The last interesting contribution to the polynomials root finding history was made by Kalantari [Kal09], who introduced the so-called polynomiography to science. Polynomiography defines

visualization process of the roots of complex polynomials approximation, using fractal and non-fractal images created via the mathematical convergence properties of iteration functions. An individual image is called a polynomiograph. Polynomiography combines both the art and science aspects. Polynomiography, as a method which generates nicely looking graphics, was patented by Kalantari in USA in 2005 [Kal09].

It is known that any complex polynomial p of degree n having n roots, according to the Fundamental Theorem of Algebra, can be uniquely defined by its coefficients $\{a_n, a_{n-1}, \dots, a_1, a_0\}$:

$$p(z) = a_n z^n + a_{n-1} z^{n-1} + \dots + a_1 z + a_0 \quad (1)$$

or by its zeros $\{z_1, z_2, \dots, z_{n-1}, z_n\}$:

$$p(z) = (z - z_1)(z - z_2) \dots (z - z_n). \quad (2)$$

Roots finding iterative process can be obviously applied to the both representations of p . As the result the polynomiographs are generated. Degree of polynomial defines the number of basins of attraction (root's basin of attraction is an area of the complex plane in which each point is convergent to the root using the root finding method). Localizations of the basins can be controlled by changing the roots positions on the complex plane manually.

Usually, polynomiographs are coloured based on the number of iterations needed to obtain the approximation of some polynomial root with a given accuracy and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

a chosen iteration method. The description of polynomiography, its theoretical background and artistic applications are described in [Kal05, Kal09].

Fractals and polynomiographs are generated by iterations. Fractals are self-similar, have complicated and non-smooth structure and are not dependent on resolution. Polynomiographs are different. Their shape can be controlled and designed in a more predictable way in opposition to typical fractals. Generally, fractals and polynomiographs belong to different classes of graphical objects.

Summing up, polynomiography can be treated as a visualization tool based on the root finding process. It has many possible applications in education, math, sciences, art and design [Kal09].

In [KGL12] the authors have used Mann and Ishikawa iterations instead of the standard Picard iteration to obtain some generalization of the Kalantari's polynomiography and have presented some polynomiographs for the cubic equation $z^3 - 1 = 0$, permutation and double stochastic matrices. Earlier, other types of iterations were used in [SJM09] for superfractals and in [PK11] for fractals generated by IFS. Also Julia sets and Mandelbrot sets [ARC14] and the so-called antifractals [RC12] were investigated using Noor iteration instead of the standard Picard iteration.

In the paper we generalize the results from [KGL12]. Thanks to the application of the new kinds of iterations we essentially extended the set of polynomiographs. Some of them are very interesting from the aesthetic point of view. They can be used as patterns for textures, in paintings creation, carpet and tapestry design etc.

The paper is organised as follows. In section 2 different types of iterations are defined. Section 3 is devoted to Newton's method for finding roots of polynomials and its generalizations, and presents some iteration formulas. In section 4 examples of polynomiographs with different types of iterations for complex equation $z^3 - 1 = 0$ are presented. The last section, section 5, describes some conclusions and plans for future work.

2 ITERATIONS

It is known that equations of the form $f(x) = 0$ can be equivalently transformed into a fixed point problem $x = T(x)$, where T is some operator [BF11]. Then, applying approximate fixed point theorem one can get information on the existence or sometimes both on existence and uniqueness of fixed point that is the solution of the starting equation.

Let (X, d) be a complete metric space and $T : X \rightarrow X$ a selfmap on X . The set $\{x^* \in X : T(x^*) = x^*\}$ is the set of all fixed points of T . In the ample literature [Ber07, Ish74, KDGE13, Kha13, Noo00, Man53, PS11, Sua05] many iterative processes have been described for the

approximation of fixed points. Below we recall some known iteration processes from the literature. Assume that each iteration process starts from any initial point $x_0 \in X$.

- The standard Picard iteration [Pic90] introduced in 1890 is defined as:

$$x_{n+1} = T(x_n), \quad n = 0, 1, 2, \dots, \quad (3)$$

- The Mann iteration [Man53] was defined in 1953 as:

$$x_{n+1} = (1 - \alpha_n)x_n + \alpha_n T(x_n), \quad n = 0, 1, 2, \dots, \quad (4)$$

where $\alpha_n \in (0, 1]$ for all $n \in \mathbb{N}$.

- The Ishikawa iteration [Ish74] was defined in 1974 as a two-step process:

$$\begin{cases} x_{n+1} = (1 - \alpha_n)x_n + \alpha_n T(y_n), \\ y_n = (1 - \beta_n)x_n + \beta_n T(x_n), \end{cases} \quad n = 0, 1, 2, \dots, \quad (5)$$

where $\alpha_n \in (0, 1]$ and $\beta_n \in [0, 1]$ for all $n \in \mathbb{N}$.

- The Noor iteration [Noo00] was defined in 2000 as a three-step process as:

$$\begin{cases} x_{n+1} = (1 - \alpha_n)x_n + \alpha_n T(y_n), \\ y_n = (1 - \beta_n)x_n + \beta_n T(z_n), \\ z_n = (1 - \gamma_n)x_n + \gamma_n T(x_n), \end{cases} \quad n = 0, 1, 2, \dots, \quad (6)$$

where $\alpha_n \in (0, 1]$ and $\beta_n, \gamma_n \in [0, 1]$ for all $n \in \mathbb{N}$.

- In 2013 Khan iteration [Kha13] was defined as the following process:

$$\begin{cases} x_{n+1} = T(y_n), \\ y_n = (1 - \alpha_n)x_n + \alpha_n T(x_n), \end{cases} \quad n = 0, 1, 2, \dots, \quad (7)$$

where $\alpha_n \in (0, 1]$ for all $n \in \mathbb{N}$.

- SP iteration [PS11] was defined in 2011 as the following three-step process:

$$\begin{cases} x_{n+1} = (1 - \alpha_n)y_n + \alpha_n T(y_n), \\ y_n = (1 - \beta_n)z_n + \beta_n T(z_n), \\ z_n = (1 - \gamma_n)x_n + \gamma_n T(x_n), \end{cases} \quad n = 0, 1, 2, \dots, \quad (8)$$

where $\alpha_n \in (0, 1]$ and $\beta_n, \gamma_n \in [0, 1]$ for all $n \in \mathbb{N}$.

- The Suantai iteration [Sua05] was defined in 2005 as a three-step iteration process with 5 parameters:

$$\begin{cases} x_{n+1} = (1 - \alpha_n - \beta_n)x_n + \alpha_n T(y_n) + \beta_n T(z_n), \\ y_n = (1 - a_n - b_n)x_n + a_n T(z_n) + b_n T(x_n), \\ z_n = (1 - \gamma_n)x_n + \gamma_n T(x_n), \end{cases} \quad n = 0, 1, 2, \dots, \quad (9)$$

where $\alpha_n, \beta_n, \gamma_n, a_n, b_n \in [0, 1]$, $\alpha_n + \beta_n \in [0, 1]$, $a_n + b_n \in [0, 1]$ for all $n \in \mathbb{N}$ and $\sum_{n=0}^{\infty} (\alpha_n + \beta_n) = \infty$.

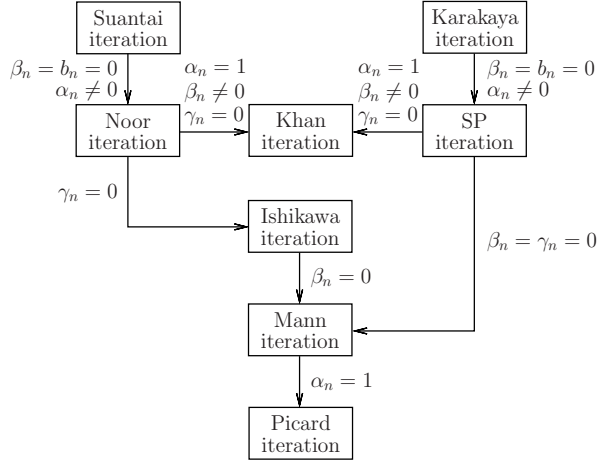


Figure 1: The diagram of iterations dependencies.

- In 2013 Karakaya et al. in [KDGE13] defined very general three-step iteration process with 5 parameters:

$$\begin{cases} x_{n+1} = (1 - \alpha_n - \beta_n)y_n + \alpha_n T(y_n) + \beta_n T(z_n), \\ y_n = (1 - a_n - b_n)z_n + a_n T(z_n) + b_n T(x_n), \\ z_n = (1 - \gamma_n)x_n + \gamma_n T(x_n), \end{cases} \quad n = 0, 1, 2, \dots, \quad (10)$$

where $\alpha_n, \beta_n, \gamma_n, a_n, b_n \in [0, 1]$, $\alpha_n + \beta_n \in [0, 1]$, $a_n + b_n \in [0, 1]$ for all $n \in \mathbb{N}$ and $\sum_{n=0}^{\infty} (\alpha_n + \beta_n) = \infty$.

The standard Picard iteration is used in the Banach Fixed Point Theorem [Ber07] to obtain the existence of the fixed point x^* of the operator T . The fixed point approximation is found under additional assumptions on the space X that it should be a Banach one and the mapping T should be contractive. The Mann [Man53], Ishikawa [Ish74] and other iterations [Ber07, KDGE13, Kha13, Noo00, PS11, Sua05] allow to weak the assumptions on the mapping T and generally allow to approximate fixed points. Dependencies between different types of iterations are presented in Fig. 1.

Our further considerations will be conducted in the space $X = \mathbb{C}$ that is obviously a Banach one. We take $z_0 \in \mathbb{C}$ and $\alpha_n = \alpha$, $\beta_n = \beta$, $\gamma_n = \gamma$, $a_n = a$, $b_n = b$ for all $n \in \mathbb{N}$ such that $\alpha \in (0, 1]$, $\beta, \gamma, a, b \in [0, 1]$, $\alpha + \beta \in [0, 1]$ and $a + b \in [0, 1]$.

3 NEWTON ROOT FINDING METHOD AND ITS GENERALIZATIONS

In this section we recall the well-known Newton method for finding roots of a complex polynomial p . The Newton procedure is given by the formula:

$$z_{n+1} = N(z_n), \quad n = 0, 1, 2, \dots, \quad (11)$$

where $N(z) = z - \frac{p(z)}{p'(z)}$, $p'(z)$ is the first derivative of p at z and $z_0 \in \mathbb{C}$ is a starting point.

Applying the Mann iteration (4) in (11) we obtain the following formula:

$$z_{n+1} = (1 - \alpha)z_n + \alpha N(z_n), \quad n = 0, 1, 2, \dots, \quad (12)$$

where $\alpha \in (0, 1]$.

Using the Ishikawa iteration (5) in (11) we get:

$$\begin{cases} z_{n+1} = (1 - \alpha)z_n + \alpha N(v_n), \\ v_n = (1 - \beta)z_n + \beta N(z_n), \end{cases} \quad n = 0, 1, 2, \dots, \quad (13)$$

where $\alpha \in (0, 1]$ and $\beta \in [0, 1]$.

Substituting the Noor iteration (6) in (11) we get:

$$\begin{cases} z_{n+1} = (1 - \alpha)z_n + \alpha N(v_n), \\ v_n = (1 - \beta)z_n + \beta N(w_n), \\ w_n = (1 - \gamma)z_n + \gamma N(z_n), \end{cases} \quad n = 0, 1, 2, \dots, \quad (14)$$

where $\alpha \in (0, 1]$ and $\beta, \gamma \in [0, 1]$.

Using the Khan iteration (7) in (11) we get:

$$\begin{cases} z_{n+1} = N(v_n), \\ v_n = (1 - \alpha)z_n + \alpha N(z_n), \end{cases} \quad n = 0, 1, 2, \dots, \quad (15)$$

where $\alpha \in (0, 1]$.

Substituting SP iteration (8) in (11) we get:

$$\begin{cases} z_{n+1} = (1 - \alpha)v_n + \alpha N(v_n), \\ v_n = (1 - \beta)w_n + \beta N(w_n), \\ w_n = (1 - \gamma)z_n + \gamma N(z_n), \end{cases} \quad n = 0, 1, 2, \dots, \quad (16)$$

where $\alpha \in (0, 1]$ and $\beta, \gamma \in [0, 1]$.

Using the Suantai iteration (9) in (11) we get:

$$\begin{cases} z_{n+1} = (1 - \alpha - \beta)z_n + \alpha N(v_n) + \beta N(w_n), \\ v_n = (1 - a - b)z_n + aN(w_n) + bN(z_n), \\ w_n = (1 - \gamma)z_n + \gamma N(z_n), \end{cases} \quad n = 0, 1, 2, \dots, \quad (17)$$

where $\alpha \in (0, 1]$, $\beta, \gamma, a, b \in [0, 1]$, $\alpha + \beta \in (0, 1]$, $a + b \in (0, 1]$.

Applying the Karakaya iteration (10) in (11) we get:

$$\begin{cases} z_{n+1} = (1 - \alpha - \beta)v_n + \alpha N(v_n) + \beta N(w_n), \\ v_n = (1 - a - b)w_n + aN(w_n) + bN(z_n), \\ w_n = (1 - \gamma)z_n + \gamma N(z_n), \end{cases} \quad n = 0, 1, 2, \dots, \quad (18)$$

where $\alpha \in (0, 1]$, $\beta, \gamma, a, b \in [0, 1]$, $\alpha + \beta \in (0, 1]$, $a + b \in (0, 1]$.

The sequence $\{z_n\}_{n=0}^{\infty}$ (or orbit of the point z_0) either converges or does not to a root of p . If the sequence converges to a root z^* then we say that z_0 is attracted to z^* . A set of all starting points z_0 for which $\{z_n\}_{n=0}^{\infty}$ converges to z^* is called the basin of attraction of z^* .

Boundaries between basins usually are fractals in nature. In [SK09] some generalizations of the classic Newton formula (11) are discussed.

All the above presented iteration processes are convergent to roots of polynomial p . Only the speed and the character of convergency is different and the basins of attraction to roots of p look differently for different kinds of iterations used. The defined iteration processes are used in the next section to obtain polynomiographs.

The application of non-standard iterations perturbs the shape of polynomial basins and makes the polynomiographs look more "fractally". The aim of using more general iterations, instead of the Picard iteration, was not to improve the speed of convergence but to create images that are interesting from aesthetic point of view.

4 EXAMPLES OF POLYNOMIOGRAPHS WITH DIFFERENT ITERATIONS

In this section we present some polynomiographs for complex polynomial equation $z^3 - 1 = 0$ using different iterations. In all examples the colour of each point in the image is determined with the help of Algorithm 1. I_q in the algorithm corresponds to the iteration processes from section 3 with a vector of the parameters $q \in \mathbb{R}^N$, where N is the number of parameters of the iteration. For the Picard iteration we use I instead of I_q .

Algorithm 1: Determination of colour

Input: $z_0 \in \mathbb{C}$ – starting point, k – maximum number of iterations, ε – accuracy, $q \in \mathbb{R}^N$ – parameters of the iteration I_q , $colours[0..k]$ – colourmap

Output: colour c of z_0

```

1  $i = 0$ 
2 while  $i \leq k$  do
3    $z_{i+1} = I_q(z_i)$ 
4   if  $|z_{i+1} - z_i| < \varepsilon$  then
5     break
6    $i = i + 1$ 
7  $c = colours[i]$ 
```

In the algorithm for a given point z_0 we iterate that point using I_q iteration process. If the modulus of the difference between two successive points in the iteration process is smaller than the given accuracy $\varepsilon > 0$ we assume that the generated sequence converge to a root of p and we stop the iteration. If we reach the maximum number of iterations k we assume that the generated sequence does not converge to any root of p . At the end we give a colour to the considered point using the iteration number at which we have left the while loop. This type of colouring is called the iteration colouring.

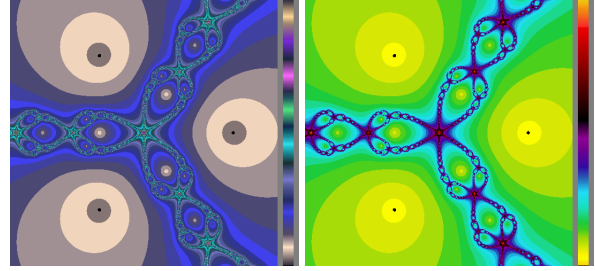


Figure 2: Picard iteration in Newton's method for $z^3 - 1$.

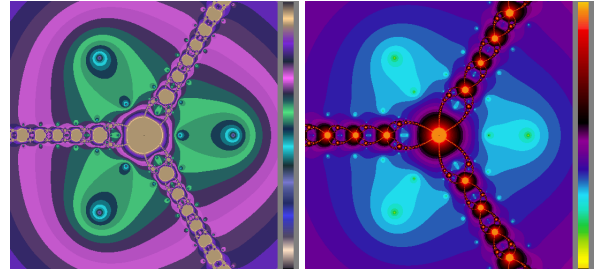


Figure 3: Mann iteration in Newton's method for $z^3 - 1$.

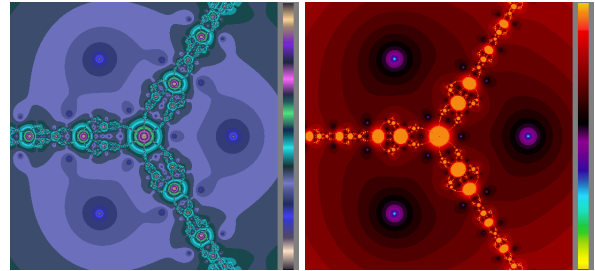


Figure 4: Ishikawa iteration in Newton's method for $z^3 - 1$.

The algorithm with different iterations from section 3 was implemented in the Processing language. Every polynomiograph in this section was computed for 1-5 seconds in average on a laptop with the Intel Core2 Duo 2 GHz CPU, 4 GB RAM.

The cubic equation $z^3 - 1 = 0$ was solved in the square domain $[-1.5, 1.5] \times [-1.5, 1.5]$ using eight different iteration processes from section 3. Images are of resolution 500×500 pixels and were generated with $k = 30$ and $\varepsilon = 0.05$ for two colourmaps. The gradient colour bar has been added to the images that shows how many iterations are needed to obtain the required accuracy ε . The performed experiments showed that values $k = 30$ and $\varepsilon = 0.05$ ensured the acceptable visual quality of polynomiographs. The obtained polynomiographs are presented in Figs. 2–9.

Generally, for polynomiographs in Figs. 2–9 one can observe that for iterations with more parameters images are more complex and are more "fractal". Polynomiographs are strongly dependent not only on iterations but also on the colourmaps used, as it can be easily seen in Fig. 10. The same graphical information con-

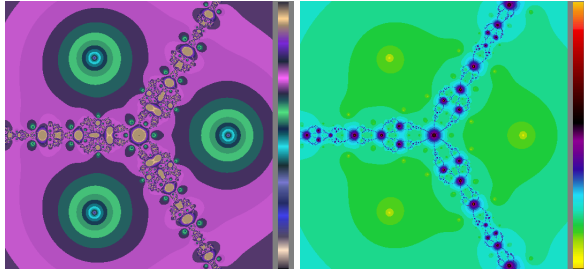


Figure 5: Noor iteration in Newton's method for $z^3 - 1$.

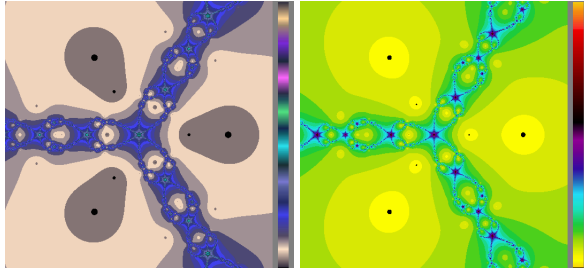


Figure 6: Khan iteration in Newton's method for $z^3 - 1$.

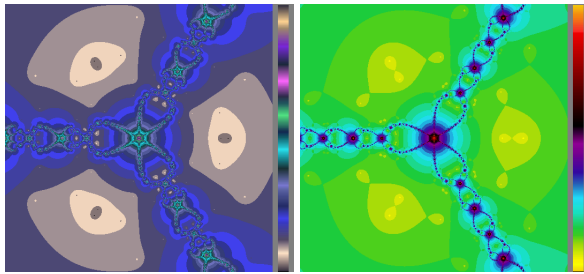


Figure 7: SP iteration in Newton's method for $z^3 - 1$.

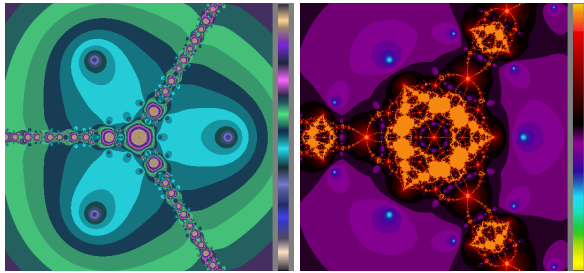


Figure 8: Suantai iteration in Newton's method for $z^3 - 1$.

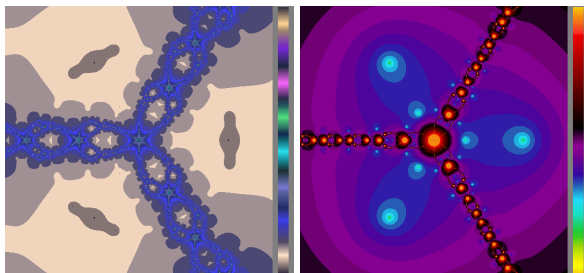


Figure 9: Karakaya iteration in Newton's method for $z^3 - 1$.

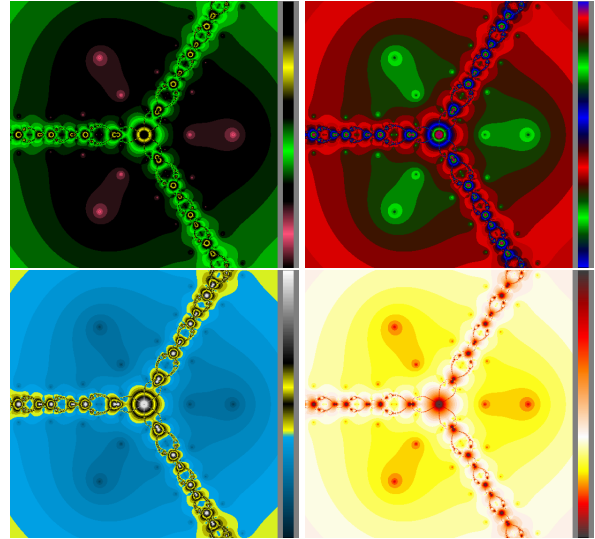


Figure 10: Fixed polynomiograph with different colourmaps.

tained in a polynomiograph may be drastically different for different colourmaps.

It should be stressed that detailed analysis of polynomiographs with respect to parameters of iterations used is difficult. But for practical use it is enough to know that if parameters are constrained and generally lie in the interval $[0, 1]$ then the iteration process is convergent and lead to images that are potentially interesting from the aesthetic point of view.

5 CONCLUSIONS AND FUTURE WORK

In the paper we presented some generalizations of the classic Newton roots finding method with non-standard iterations and some corresponding polynomiographs for exemplary complex polynomial equation $z^3 - 1 = 0$. By changing parameters of different iteration processes one can obtain a huge collection of polynomiographs essentially richer in comparison to the case of the standard Picard iteration. Further generalizations can be obtained with the help of higher order methods based on the Basic and the Euler-Schröder families of iterations [Kal09, Kal11] with non-standard iteration processes. The use of complex valued parameters instead of the real ones, as was checked by the authors, and the use of various convergence tests, as in [Gda13], lead to nice modifications of polynomiographs, as well. Additionally, different colourmaps have great influence on the aesthetic appearance of the polynomiographs.

The above mentioned problems determine our further investigations. We believe that the results of the paper can be interesting to those whose work or hobbies are related to automatically created nicely looking images. In our opinion non-standard iterations can be applied to

increase functionality of the existing polynomiography software, as well.

6 REFERENCES

- [ARC14] Ashish, Rani, M., Chugh, R.: Julia Sets and Mandelbrot Sets in Noor Orbit. *Applied Mathematics and Computation* 228(1), 615-631, (2014)
- [Ber07] Berinde, V.: *Iterative Approximation of Fixed Points*, 2nd edn. Springer, Heidelberg, (2007)
- [BF11] Burden, R.L., Faires, J.D.: *Numerical Analysis*, 9th edn. Brooks-Cole, Boston, (2011)
- [Gda13] Gdawiec, K.: Polynomiography and Various Convergence Tests. *WSCG Communication Proceedings*, pp. 15-20, (2013)
- [Ish74] Ishikawa, S.: Fixed Points by a New Iteration Method. *Proceedings of the American Mathematical Society* 44(1), 147-150, (1974)
- [Kal05] Kalantari, B.: Polynomiography: From the Fundamental Theorem of Algebra to Art. *Leonardo* 38(3), 233-238, (2005)
- [Kal09] Kalantari, B.: *Polynomial Root-Finding and Polynomiography*. World Scientific, Singapore, (2009)
- [Kal11] Kalantari, B.: Alternating Sign Matrices and Polynomiography. *The Electronic Journal of Combinatorics* 18(2), 1-22, (2011)
- [KDGE13] Karakaya, V., Doğan, K., Gürsoy, F., Ertürk, M.: Fixed Point of a New Three-Step Iteration Algorithm under Contractive-like Operators over Normed Spaces. *Abstract and Applied Analysis*, vol. 2013, Article ID 560258, 9 pages, (2013)
- [Kha13] Khan, S.H.: A Picard-Mann Hybrid Iterative Process. *Fixed Point Theory and Applications*, vol. 2013, Article ID 69, 10 pages, (2013)
- [KGL12] Kotarski, W., Gdawiec, K., Lisowska, A.: Polynomiography via Ishikawa and Mann Iterations. In: G. Bebis et al. (eds.) *Advances in Visual Computing, Part I. LNCS 7431*. Springer, Berlin, pp. 305-313, (2012)
- [Man83] Mandelbrot, B.: *The Fractal Geometry of Nature*. W.H. Freeman and Company, New York (1983)
- [Man53] Mann, W.R.: Mean Value Methods in Iteration. *Proceedings of the American Mathematical Society* 4(3), 506-510, (1953)
- [Noo00] Noor, M.A.: New Approximation Schemes for General Variational Inequalities. *Journal of Mathematical Analysis and Applications*, 251(1), 217-229, (2000)
- [PS11] Phuengrattana, W., Suantai, S.: On the Rate of Convergence of Mann, Ishikawa, Noor and SP Iterations for Continuous Functions on an Arbitrary Interval. *Journal of Computational and Applied Mathematics* 235(9), 3006-3014, (2011)
- [Pic90] Picard, E.: Mémoire sur la théorie des équations aux dérivées partielles et la méthode des approximations successives. *Journal de Mathématiques Pures et Appliquées* 6(4), 145-210, (1890)
- [PK11] Prasad, B., Katiyar, B.: Fractals via Ishikawa Iteration. In: P. Balasubramaniam (ed.) *Control, Computation and Information Systems. CCIS 140*. Springer-Verlag, Berlin Heidelberg, pp. 197-203, (2011)
- [RC12] Rani, M., Chugh, R.: Dynamics of Antifractals in Noor Orbit. *International Journal of Computer Applications* 57(4), 11-15, (2012)
- [SJM09] Singh, S.L., Jain, S., Mishra, S.N.: A New Approach to Superfractals. *Chaos, Solitons and Fractals* 42(5), 3110-3120, (2009)
- [Sua05] Suantai, S.: Weak and Strong Convergence Criteria of Noor Iteration for Asymptotically Non-expansive Mappings. *Journal of Mathematical Analysis and Applications* 311(2), 506-517, (2005)
- [SK09] Susanto, H., Karjanto, N.: Newton's Method's Basins of Attraction Revisited. *Applied Mathematics and Computation* 215(3), 1084-1090, (2009)

Spatter Tracking in Laser- and Manual Arc Welding with Sensor-level Pre-processing

Olli Lahdenoja
Technology Research Center
Brahea Center, University of
Turku, 20014 Turku Finland
olanla@utu.fi

Tero Sääntti
Technology Research Center
Brahea Center, University of
Turku, 20014 Turku Finland
teansa@utu.fi

Mika Laiho, Jonne Poikonen
Technology Research Center
Brahea Center, University of
Turku, 20014 Turku Finland
mlaiho@utu.fi, jokapo@utu.fi

ABSTRACT

This paper presents methods for automated visual tracking of spatters in laser- and manual arc welding. Imaging of the welding process is challenging due to extreme conditions of high radiated light intensity variation. The formation and the number of spatters ejected in the welding process are dependent on the parameters of the welding process, and can potentially be used to tune the process towards better quality (either on-line or off-line). In our case, the spatter segmentation is based either on test on object elongatedness or Hough transform, which are applied on pre-processed image sequences captured by a high-speed smart camera. Part of the segmentation process (adaptive image capture and edge extraction) is performed on the camera, while the other parts of the algorithm are performed off-line in Matlab. However, our intention is to move the computation towards the camera or an attached FPGA.

Keywords

Laser welding, Thick Steel, Metal arc welding, Spatter tracking, Smart camera.

1. INTRODUCTION

One phenomenon in laser welding and welding in general is the formation of spatters, which represent the ejection of melt from the interaction zone caused by extreme conditions of material interaction. Increased spatter formation may indicate that the process parameters need to be adjusted to improve the quality.

In previous works [Nic11a], [Nic09a] the spatter formation was related to the on-line control of process parameters in welding of aluminum and 1mm thick steel sheets. In [Nic11a] the spatter segmentation was based on blob analysis including a threshold operation with a suitable mask. In [Jäg08a] a spatter tracking system incorporating a Kalman filter for trajectory analysis was proposed. In [Fen09a] spatter tracking in the context of laser MAG (metal active gas) welding was proposed. An off-line Kalman filter based approach for tracking

was implemented with an SVM based spatter extraction.

In this paper, we present an approach for spatter tracking in the scenario of welding thick steel, by a combination of a smart camera for on-line processing and spatter event tracking as an off-line process, enabling automated analysis of the spatter behavior in welding. Different alternatives for spatter segmentation and tracking are considered in the context of high power laser welding and traditional manual metal-arc welding.

2. SPATTER EXTRACTION AND SEGMENTATION

On-line analysis

The huge difference between the intensity of light radiated from the welding zone compared to the surrounding areas causes a challenge for welding spatter segmentation and tracking. The objective is to enable the spatter object segmentation both within the very low and the very high intensity regions. This can be difficult to achieve because of the high intra-scene dynamic variation. Successful image capture may still be performed with the aid of active laser illumination [Fen09a]. Alternatively, a smart camera adapting to the intensity variation locally can be used.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Our approach is to use a high-speed smart camera, which enables the imaging of the welding zone by adjusting the integration time of the individual pixels according to their nearby local surroundings [Lah13a]. After this, the camera is used to extract the edge information from the second order spatial neighbors against an adjustable threshold. The KOVA1 array [KOV] provides a 96x96 resolution pixel parallel mixed-mode imaging processor, with a processor per pixel image processing functionality. In order to achieve a high enough frame rate for this application, only binary (1-bit) image sequence was used. Using gray scale images would limit the frame rate by filling up the bandwidth of the 100 Mb/s Ethernet connection from the camera to the PC. Additionally, the required processing (edge detection) is performed faster on the camera than on the FPGA or the PC.

Off-line analysis

The spatter segmentation was performed as an off-line process in Matlab. Two slightly different methods for spatter segmentation are considered. The first is to apply Hough transform directly on the extracted edge images. This allows capturing the center location and perimeter of the spatters. The disadvantage of the method is that spatters which are not circular cannot be detected correctly. However, the method can also be implemented on FPGA for high-speed analysis.

The second considered approach is first to apply an additional morphological processing step to the extracted binary images (a morphological closing), in combination with connected component labeling. The closing operation could also be implemented on-line. The extracted object regions are then tested by their eccentricity ('regionprops' in Matlab R2012b). An additional step considered in this paper is to require that the center pixel of the object is of adjacent polarity compared to the edge pixels. This may be used to refine the circularity of the object, since it can be assumed that the spatters are, in general, symmetric and the edge image provided by the adaptive integration of the camera provides an object with a hole in the center.

3. SPATTER TRACKING

The proposed tracking algorithm is implemented in Matlab, and described in the following. First, the segmented objects are extracted as described in the previous section. A list containing the elements of initial spatter properties (SPL1) is created. Each spatter is assigned an individual ID number (in phase 3) and spatter area property, which describes a circular region in the case of Hough transform, and

otherwise a rectangular region. TSPL_PRE (previous tracked spatter list) is initialized also to zero in this step. The steps of the tracking algorithm are explained below (1-5).

- 1) Initialize TSPL = 0 (Tracked spatter list) and read SPL2 (Spatter list 2) based segmented spatters at this frame.
- 2) Compare SPL2 to TSPL_PRE according to criterion IS_NEAR and update the matching spatters to TSPL. Remove these spatters from SPL2.
- 3) Compare the remaining SPL2 to SPL1, initialize and update those with property IS_NEAR_INIT to TSPL. Initialize spatter ID.
- 4) SPL1=SPL2, TSPL_PRE=TSPL
- 5) If frames left, Go back to phase 1

To initiate the tracking, the test IS_NEAR_INIT considers the spatter locations and areas only, without considering the previous displacement vector. In this phase, we use Euclidean maximum distance requirement of below 15 pixels. The test IS_NEAR performs the following action. The previous displacement vector (See Fig. 1) of the spatter under tracking has been stored and it is compared by using the IS_NEAR criteria to each of the new spatter observations. We use a Euclidean distance requirement of < 5 pixels. The assumption is that the spatters roughly propagate with constant velocity. An additional property which is checked is that the size of the new spatter has to be above 0.5 times and below 2 times of the size of the old spatter. The time which the spatter has been tracked (in number of frames) is also stored. All of the tracked spatters are collected to a list for later analysis. The post-processing step incorporates also a possibility to consider the average sizes and velocities of the spatters under test.

Observe that we use the same tracking algorithm in both cases i.e. with the connected component labeling based segmentation and with the Hough transform based segmentation. Thus, the segmentation algorithm, which could be performed on the camera or on an FPGA, is not affected by the tracking algorithm. Further possibilities of HW accelerating the tracking algorithm are discussed in section 6.

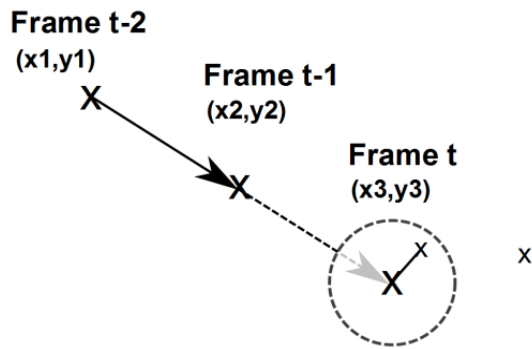


Figure 1. The test for proximity between two spatters. The estimated new center location of the spatter according to previous frames is compared to each of the candidate spatters in Frame t according to Euclidean distance (partly reminding the Kalman filter [Jäg08a]).

4. EXPERIMENTAL SETUP

Manual metal-arc welding set-up

The first tests were performed with manual arc welding, where the number of spatters is very large. The experimental setup in this case was initially targeted to test the properties of the camera, but the excessive spatter formation was shown to be suitable also for testing the spatter tracking algorithm. The frame rate used was 1408 Fr/s (with a difference between two frames of approximately 710 μ s). Fig. 2 represents an example of the adaptive camera capture and the corresponding edge image captured on-line.

Laser welding set-up

The second test case used was a coaxial setup of full-penetration mode laser welding of 4mm thick steel. The laser power applied was selected so that full penetration was achieved, 4.5kW on 12mm/s and 6kW on 20mm/s welding speeds. The number of spatters was very low, which created challenges for the automated spatter analysis due to a relatively large portion of vapor plume. A partial penetration mode is expected to produce more spatters, which will be studied in the future. The frame-rate applied was approximately 3500 Fr/s.

5. EXPERIMENTS

Manual metal-arc welding tests

Hough transform was found to suite well to the case of manual metal arc welding, since the spatters were typically circular (See Fig. 2). The considered radiuses were from 2 to 10, with a Hough sensitivity parameter (HS) of 0.85 (the available implementation

in Matlab R2012b). However, large spatters (e.g. $r > 6$) were very rare, due to larger imaging area in comparison with the laser welding. Connected component labeling, which was applied after a binary closing with a 3x3 mask of all ones was used as a reference method. The object elongatedness was required in this case to be below 0.9.

Since it was not possible to manually label the whole welding sequence, the performance of the tracking was estimated so that a ground truth was generated and the overall spatter count returned by the algorithm (i.e. TP+FP) was determined at constant 100 frames steps (see Fig. 4, index 1 indicates frames 1-100 etc.). Note also that the ground truth contains all spatters, but the tracking algorithm only counts spatters included in at least 3 succeeding frames.

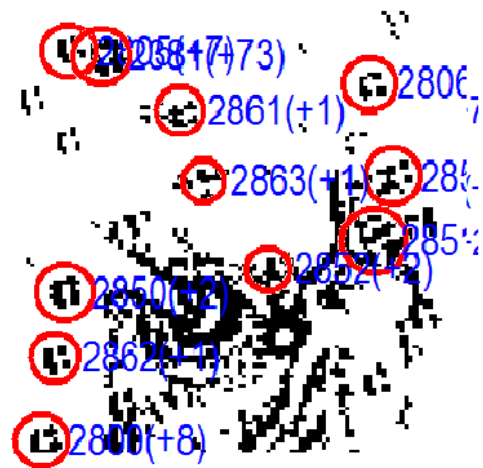
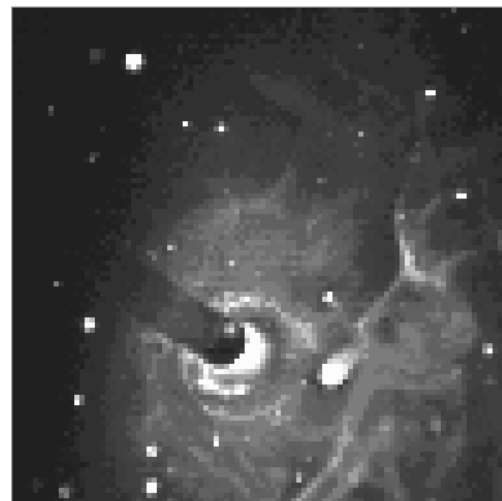


Figure 2. The arc welding set-up. The image on the top represents the welding zone captured by the adaptive integration. On the bottom the edges extracted on-line and the tracked spatter IDs extracted off-line.

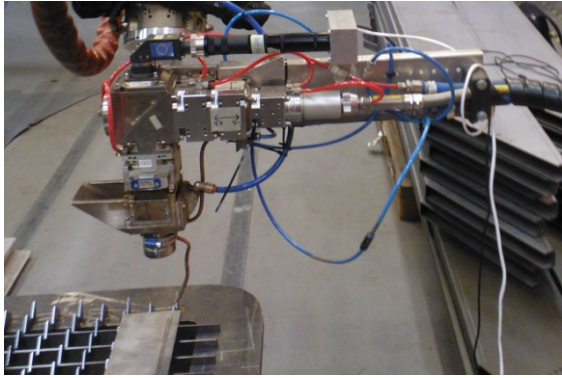


Figure 3. The laser welding test setup. The camera is attached to the welding head (from Precitec) through a 90 degree mirror and focusing lens. The fibre laser is brought from a collimator through 90 degree mirror optics to the work piece. The focal position of the laser was -4mm (below the surface).

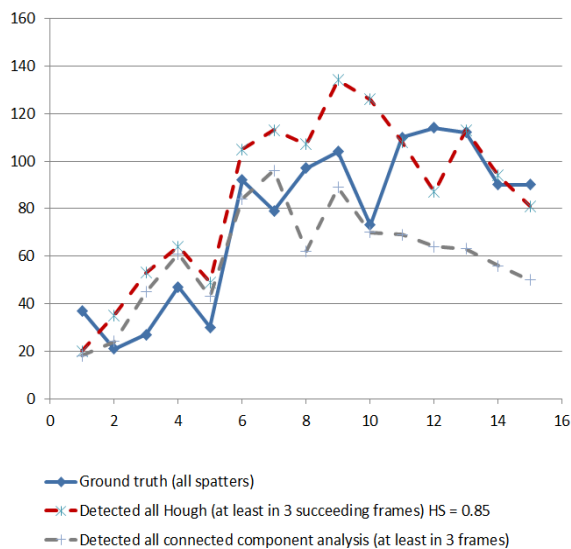


Figure 4. An indicative chart on the ground truth and the overall number of detected spatters in arc welding in frames 1-1500 at 100 frame intervals.

In the tests, increasing the requirement of the number of consecutive tracked frames could be used to filter out false positives caused by the vapor plume and to make it possible to increase the sensitivity of the segmentation. Also, a rectangular area near to the arc end was masked to reduce the effect of the plume. The ground truth was generated by three human experts in 500 frame intervals, which was quite challenging and could induce certain bias. It can be observed, that results follows the manually extracted spatter behavior. It appears in Fig.4, that when the number of spatters (and likely the amount of plume) is higher, the Hough transform gives better results. In Fig. 5, an angular histogram plot of the spatter directions for the same sequence is shown with

Hough transform (3 succeeding tracked frames required).

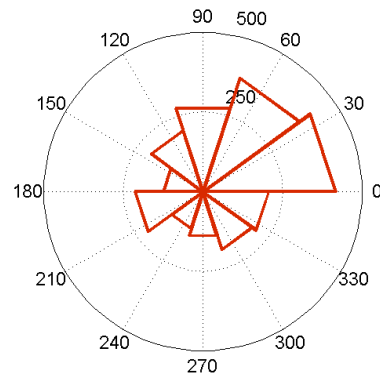


Figure 5. An angular histogram plot of the most common spatter directions in the arc welding test using Hough transform. The location of the arc was approximately in the location shown in Fig. 2.

Laser welding tests

Since the spatters in laser welding were not always circular, the connected component analysis in combination with spatter elongatedness test was applied with the same parameters as in arc welding. An additional requirement was that the center location of the spatters was white, which was shown to reduce the false alarm rate. The size of the detected spatters was limited to a square of minimum edge dimension of 5 pixels and maximum edge dimension of 20 pixels.

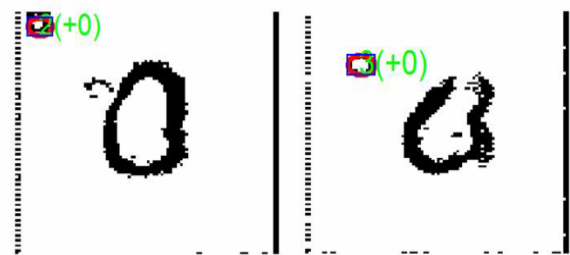


Figure 6. Examples of tracked spatters in laser welding scenario.

A ground truth of first 1000 frames of the sequence was manually collected, including only 13 spatters. The total amount of spatters was very low due to the used full penetration mode. A total of 9 spatters were detected correctly in comparison with the ground truth with one false alarm. Thus, a detection rate of 69.2% was obtained with false positive rate of 0.001/frame. Fig. 6 represents examples of tracked spatters in the laser welding scenario.

In [Nic11a] the false positive rates of spatter segmentation varied between 0.009/frame and 0.028/frame. The detection rate is not directly comparable, since the authors in [Nic11a] consider the detection of spatter events as an occurrence of one or more spatter within a single image. If there were one or more spatters present in the ground truth and also in the segmentation result, the match was considered as positive. In our case, individual spatters, which may locate across multiple succeeding frames are considered. However, it should be noted, that the tracking option used in our paper relaxes the requirement for spatter counting, since if a spatter is not found from an individual frame, the correct detection can still be attained in the succeeding frames.

6. ON POSSIBILITIES OF HW ACCELERATING THE ALGORITHMS

We are currently studying options for implementing the segmentation and the tracking algorithm on FPGA or by in-camera processing. While examining seam tracking, we have already demonstrated Hough transform based line detection on the FPGA attached to the camera [Sän14a]. The unit will be modified to detect circles, and then used as the detection engine for the tracking algorithm presented in this paper. Also the FPGA based acceleration of the tracking algorithm is possible. Currently on laptop PC with Matlab the algorithm runs approximately 30fps (without considering the time for the segmentation).

FPGA based processing

In order to implement the tracking algorithm on an FPGA the number of segmented spatters (elements in the lists) sets the limit for the efficient implementation. This is because the parameters of the spatter elements (e.g. 8-bit values) should be, preferably, mapped to the internal FPGA Block RAMs and not to external memory, which introduces extra latency. Assuming a BRAM size of 36kbit and list element size of $6 \times 8 = 48$ bits, the maximum number of elements in a list can be estimated. When mapped to a single BRAM the maximum length of a list would be 768, which should be enough for the application considered. The collection of the final spatter properties for further analysis could then be implemented by sending the data to a PC. At this point, as mentioned, only the adaptive image capture and spatter edge extraction were performed on-line.

Camera based processing

The applied smart camera (KOVA1) includes pixel-level processing functionality based on Cellular Nonlinear Network (CNN)-type [Ros93a] parallel

processing architecture. In the KOVA1 CMOS sensor chip, analog and digital (mixed-mode) computational circuitry, as well as binary (static) and analog (dynamic) memory is integrated physically together with each sensor diode, creating a focal plane cell [Lai11a]. The additional pixel-level circuitry naturally imposes some limits on practical image resolution vs. chip size (cost/yield), e.g. the size of the commercial KOVA1 sensor-processor array is 96x96 pixels, while the largest published research implementation with a similar cellular processing architecture is the SCAMP-5 with a 256x256 pixel array [Car13a].

Each pixel cell in the sensor array of the KOVA1 is directly connected to its nearest neighborhood and in some operations even to the second neighborhood. Information propagation over larger distances is also possible in an asynchronous or synchronous manner. The cellular connectivity together with local, pixel-level, memory and processing circuitry allows the implementation of multiple-stage image analysis operations on the pixel plane. The pixel-level parallel processing and the avoidance of data transfer to an external processor result in reduced energy consumption and more importantly considering the present application, greatly reduced processing delay. The pixel-level processing functionality of the KOVA1 includes both grayscale and binary (1 bit) operations. The grayscale processing was used in this case to implement the image dynamic compression used in the capture of the welding data. The preprocessed grayscale image is segmented into a 1 bit binary representation to allow more efficient pixel-level shape and object analysis.

At this stage an edge detection operation has been used to extract relevant detail. The pixel level processing circuitry enables the implementation of sensor-level binary mathematical morphology and other similar analysis, as described in [Lai08a], as well as normal Boolean logic operations. By storing and combining intermediate result images in local pixel memory even complex spatial processing sequences can be performed on the sensor plane, including time-domain inter-frame analysis. Asynchronous binary propagation on the sensor plane can be used to perform regional operations such as holefilling, or morphological reconstruction very efficiently. By applying more binary analysis on the sensor-level, the processing demands at later stages (e.g. on an FPGA) can be greatly reduced, while retaining sufficient overall analysis speed.

If edge image based spatter segmentation is applied, one possibility to perform more computation at the camera level would be to use a holefiller template with pixel-level processing to detect roundish spatter objects, and then to extract the center locations of these objects for further processing. However, the

holefiller template requires, that the edges of the objects are fully connected, which may not be always achievable. This could be avoided by using orientation selective propagation templates and by combining them or by applying a combination of dilation and skeleton extraction. Also blob based region extraction could be considered in the future.

7. DISCUSSION

The test set used in laser welding was challenging, since the spatters were not always circular, and due to second order neighborhood used in smart camera based segmentation. A second order neighborhood was used to better obtain the smooth edges of the spatters. Due to this, the edges of the spatters were thick, thus reducing the freedom in applying a larger closing mask. The purpose of the closing mask is to avoid a situation, where an individual spatter is detected as two distinct objects in the connected component labeling step. Also, if using too large closing mask, the centroid of the spatters becomes filled, and the segmentation process becomes less robust.

An approach for spatter segmentation and tracking in two welding scenarios was proposed. Our approach extends the work in [Nic11a] in that it also enables the spatter tracking and the analysis of the individual spatter properties. Future work consists of developing methods towards on-line spatter analysis. In comparison with work in [Lah13a] the methods proposed provide further robustness to false positives induced by the vapor plume. A limitation of the proposed approach in comparison with traditional multi object tracking is that the collision of two spatters may lead to erroneous tracking. This issue should be addressed in the future.

8. CONCLUSIONS

Methods for spatter segmentation and tracking in two welding scenarios were proposed in this paper. Hough transform was proposed for extracting spatters in the manual-arc welding scenario, while object elongatedness analysis was proposed for laser welding. The advantage of the Hough transform based segmentation is that it inherently avoids merging two distinct objects (or plume) nearby into the same object in the segmentation phase. The main advantage of the connected component analysis approach and the elongatedness test was that the objects need not to be circular, allowing flexibility for segmentation. Checking the polarity of the center pixel of the extracted object also increased the robustness of the segmentation algorithm.

9. ACKNOWLEDGMENTS

We would like to acknowledge the support from Lappeenranta University of Technology (LUT) and Machine Technology Center Turku in conducting the tests. The research was funded by Academy of Finland project no. 254430.

10. REFERENCES

- [Nic11a] Nicolosi, L., Tetzlaff, R., Blug, A., et.al, A Monitoring System for Laser Beam Welding Based on an Algorithm for Spatter Detection. Conf.proc. ECCTD'11, pp. 25-28.
- [Nic09a] Nicolosi, L., Tetzlaff, R., F. Abt, et. al, New CNN based algorithms for the full penetration hole extraction in laser welding processes: Experimental results, IJCNN'09, pp. 2256-2263.
- [Jäg08a] Jäger, M., Humbert, S., Harmrecht, F. A., Sputter Tracking for the Automatic Monitoring of Industrial Laser-Welding Processes. IEEE Trans. Industrial Electronics, 55, No.5, 2008.
- [Fen09a] Fennander, H., Kyrki, V., Fellman, A., Salminen, A., Kälviäinen, H., Visual measurement and tracking in laser hybrid welding, Machine Vision and Applications 20, pp. 103–118, (2009).
- [Lah13a] Lahdenoja, O., Sääntti, T., Poikonen, J., Laiho, M., Paasio, A., Characterizing Spatters in Laser Welding of Thick Steel Using Motion Flow Analysis, SCIA'13, pp. 675-686.
- [KOV] <http://www.kovilta.fi>
- [Sän14a] Sääntti, T., Lahdenoja, O., Paasio, A., Laiho, M., Poikonen, J., Line Detection on FPGA with parallel sensor-level segmentation, CNNA'14, to Appear.
- [Ros93a] Roska, T., Chua, L. O., The CNN Universal Machine: An Analog Array Computer, IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing. 40(3), pp. 163-173, (1993)
- [Lai11a] Laiho, M., Poikonen J., Paasio A.: MIPA4k: Mixed-Mode Cellular Processor Array. In: Zarandy A. (ed.): Focal-Plane Sensor-Processor Chips, Springer (2011)
- [Car13a] Carey, S.J., Lopich, A., Barr, D.R.W., Wang, Bin, Dudek, P., A 100,000 fps vision sensor with embedded 535GOPS/W 256×256 SIMD processor array, Symposium on VLSI Circuits (VLSIC), pp. 182-183, (2013)
- [Lai08a] Laiho, M., Paasio, A., Flak, J., et. al, Template Design for Cellular Nonlinear Networks With 1-Bit Weights. IEEE Trans. on Circuits and Systems 55(3), pp. 904-913 (2008)

Fast and Robust Tessellation-Based Silhouette Shadows

Tomáš Milet
Brno University of
Technology Czech
Republic
imilet@fit.vutbr.cz

Jozef Kobretek
Brno University of
Technology Czech
Republic
ikobretek@fit.vutbr.cz

Pavel Zemčík
Brno University of
Technology Czech
Republic
zemcik@fit.vutbr.cz

Jan Pečiva
Brno University of
Technology Czech
Republic
peciva@fit.vutbr.cz

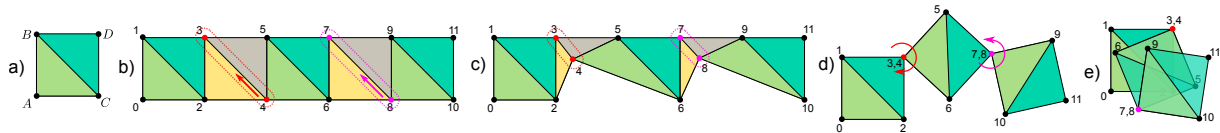


Figure 1: The image shows the transformation of a quad into three overlapping shadow volume sides. The transition from part a) to part b) is tessellation of quad with Multiplicity = 3. Only green and blue triangles will be drawn. Yellow and gray triangles will be degenerated. The transition from part b) over part c) to part d) shows degeneration process. Red and purple vertices 3, 4 and 7, 8 from part a) form only one vertex in part d). The transition from part d) to part e) shows rotation around red and purple vertices. This transformation creates three overlapping sides of shadow volume. Positions of vertices A, B, C, D that form initial quad, can be computed according to equations 2.

ABSTRACT

This paper presents a new simple, fast and robust approach in computation of per-sample precise shadows. The method uses tessellation shaders for computation of silhouettes on arbitrary triangle soup. We were able to reach robustness by our previously published algorithm using deterministic shadow volume computation. We also propose a new simplification of the silhouette computation by introducing reference edge testing. Our new method was compared with other methods and evaluated on multiple hardware platforms and different scenes, providing better performance than current state-of-the-art algorithms. Finally, conclusions are drawn and the future work is outlined.

Keywords: shadows, shadow volumes, silhouette, tessellation shaders, geometry shader

1 INTRODUCTION

Shadow Volumes (SV) algorithm was introduced in 1977 by [Crow, 1977], first implementation using hardware support via stencil buffer was carried out by [Heidmann, 1991]. Heidman's implementation is generally called z-pass, but does not produce correct results when observer is in shadow. This problem was eliminated in the z-fail method [Everitt and Kilgard, 2002], which reverses depth test function, but requires shadow volumes to be capped.

Shadow Mapping (SM) algorithm, proposed by [Williams, 1978], is an alternative approach to shadow volumes. It uses depth information from light source stored in a texture. Shadow mapping is nowadays massively used in games thanks to its performance,

but suffers from spatial and often temporal aliasing problems and produces imperfect shadows because of limited shadow map resolution [Donnelly and Lauritzen, 2006]. Low resolution is not an issue for games, because scenes can be adjusted so that visual artifacts are suppressed or a filtering method is applied, but applications for visualization in architecture or industrial design require pixel-correct shadows for object visualization. Per-sample precise alias-free shadow maps (AFSM) algorithm was proposed by [Sintorn et al., 2008]. Their method stores multiple samples into a list for each shadow map pixel and conservatively rasterizes triangles into shadow map using CUDA. Individual samples stored in lists are then tested against shadow volume of the triangle. As they stated, their per-sample precise method is three times slower than standard shadow mapping with resolution of 8096 by 8096 texels.

While producing per-sample correct shadows, SV are affected by performance issues. In its naive form, when a volume is generated for every triangle in the scene, resulting performance is very low due to rasterization of a large amount of triangles. More efficient way is to construct shadow volumes only from silhouette edges

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

of the occluding geometry, which has positive impact on fill-rate. Silhouette extraction on CPU was first published by [Brabec and Seidel, 2003].

Several silhouette-based methods were published since then, utilising novel hardware features to speed up silhouette calculation. [McGuire et al., 2003] managed to implement the algorithm in vertex shader and [Stich et al., 2007] used geometry shaders.

Most of the methods mentioned above are not completely robust and also cannot handle non 2-manifold casters. [Kim et al., 2008] proposed an algorithm for non 2-manifold casters, but unfortunately it is not robust. Kim's algorithm was improved in [Pečiva et al., 2013] using deterministic multiplicity calculation, which we further simplified in this paper.

[Sintorn et al., 2011] also proposed a shadowing technique based on CUDA software rasterization of per-triangle shadow frusta. This technique uses a small bias when testing sample depth against a triangle plane to avoid self-shadowing. This bias, however, may cause a shadowed fragment to be lit in the final result, moreover, it is also scene-dependent.

2 METHOD DESCRIPTION

We have developed three methods - two per-triangle approaches and robust silhouette method.

Our silhouette method is based on the work of [Kim et al., 2008]. This algorithm calculates so-called *multiplicity* of an edge - light plane from light source through the edge is casted and all opposing vertices are tested, if they are above or below the plane. According to result of the test, multiplicity is incremented or decremented. Absolute value of multiplicity is the number of times an infinite quad needs to be drawn from this edge.

2.1 Per-Triangle Methods

These methods require no pre-processing and work with arbitrary triangle soup. In the first variant, input patch has 3 points, which are original points of the triangle. Tessellation factors are 3 (inner) and 1 (outer, for all sides), equal spacing and reversed triangle winding. The resulting patch can be seen in the picture 2b.

We construct a simple volume in evaluation shader as in Algorithm 1. Front cap needs to be rendered in second pass in order to close the volume.

We also designed a single-pass version for z-fail. This method takes a triangle as an input, but adds one more point to form a quad (4 control shader invocations per patch). This quad is then tessellated using outer factors (1, 5, 1, 5), inner (5,1) and fractional odd spacing, resulting in a shape seen in Fig. 3b.

Evaluation shader then twists the shape in order to create a volume, note Figure 3.

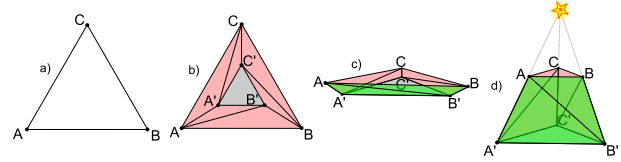


Figure 2: Creating semi-enclosed shadow volume from a triangle. Initial triangle in a) is tessellated using outer factors (1, 1, 1) and inner (3) b). Points A' , B' , C' are given positions of points A , B , C c) and then pushed to infinity to form a volume with back cap d).

Data: original points $\mathbf{P}[3]$, light position \mathbf{L} , tessellation coordinates $\mathbf{T} = (x, y, z), x, y, z \in \langle 0, 1 \rangle$

Result: world-space coordinates X

$c = x \cdot y \cdot z;$

if $c == 0$ **then**

$\mathbf{X} = \mathbf{P}[0] \cdot x + \mathbf{P}[1] \cdot y + \mathbf{P}[2] \cdot z;$

$\mathbf{X}_w = 1;$

else

$i = \text{getIndexOfLargestVectorElement}(\mathbf{T});$

$\mathbf{X} = l_w \cdot \mathbf{P}[i] - \mathbf{L};$

$\mathbf{X}_w = 0;$

end

Algorithm 1: Evaluation shader in two-pass per-triangle method

2.2 Silhouette Method

The method finds silhouette edges by looping over every edge in the model. Each edge is processed in parallel in Tessellation Control Shader where multiplicity is computed. An input patch primitive is composed of two vertices that describe an edge, one integer that contains number of opposite vertices and n opposite vertices, see Figure 4. Because patch the size must be constant, some positions are not used.

A vertex buffer of model has to be extended by E_n vertices, which is the number of edges in the model. We used element buffer to reduce memory requirements.

Byungmoon's algorithm [Kim et al., 2008], as in its core proposal, has a flaw that multiplicity is not calculated in a deterministic way. In older approach [Pečiva et al., 2013], it was fixed by calculating multiplicity per triangle and if the 3 results throughout all 3 edges were not consistent, we discarded the triangle from further processing, because it meant that the triangle is almost parallel to the light and does not cast a shadow. We further improved this approach - multiplicity is now computed only once for each opposite vertex using *reference edge*.

A choice of reference edge has to be the same for all occurrences of a triangle. This can be achieved for ex-

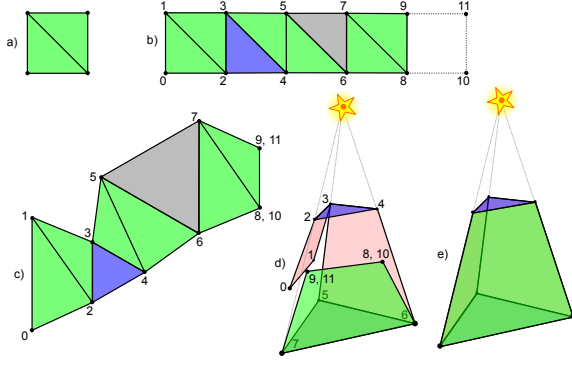


Figure 3: Single-pass per triangle method, a full shadow volume is created in a single pass. One point is added to the triangle in order to form a quad *a*) which is then tessellated using factors (1, 5, 1, 5), (5, 1) *b*). Points 10 and 11 are merged with 8, 9. Light cap is visualized as blue, dark cap grey *c*). Then we join points 0-7, 1-5, 2-9, 4-8 and push points 5, 6, 7 to infinity *d*) to make the volume *e*).

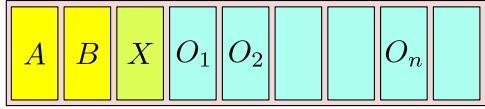


Figure 4: Input patch for tessellation control shader

ample by introducing vertex ordering - Equations 1 and Algorithm 2.

$$\begin{aligned} A < B &\Leftrightarrow \text{Greater}(A, B) < 0 \\ A = B &\Leftrightarrow \text{Greater}(A, B) = 0 \\ A > B &\Leftrightarrow \text{Greater}(A, B) > 0 \end{aligned} \quad (1)$$

Data: Vertices A, B

Result: Result r of comparison

$S = \text{sgn}(A - B)$;

$K = (4, 2, 1)$;

$r = S \cdot K$

Algorithm 2: Function $\text{Greater}(A, B)$ used for vertex ordering.

In order to guarantee consistency, reference edge of a triangle in our algorithm is constructed using smallest and largest vertex of a triangle, as in Algorithm 2. More options for such method are available, but evaluation per each triangle occurrence must be consistent in order to get correct results.

To simulate behaviour of Byungmoon's algorithm (edge casts a quad as many times as it has multiplicity), we tessellate the casted quad from the edge using inner tessellation levels ($\text{Multiplicity} \cdot 2 - 1, 1$) and then we bend the tessellated quad in evaluation shader in a way to create m overlapping quads, as seen in Fig. 1, which demonstrates edge $A-B$ having multiplicity of 3.

The procedure of multiplicity calculation is described in Algorithm 3 and 4.

Data: Edge $A, B, A < B$, set \mathcal{O} of opposite vertices $O_i \in \mathcal{O}$, light position L in homogeneous coordinates

Result: Multiplicity m

$m = 0$;

for $O_i \in \mathcal{O}$ **do**

if $A > O_i$ **then**

$m = m + \text{CompMultiplicity}(O_i, A, B, L)$;

else

if $B > O_i$ **then**

$m = m - \text{CompMultiplicity}(A, O_i, B, L)$;

else

$m = m + \text{CompMultiplicity}(A, B, O_i, L)$;

end

end

end

Algorithm 3: Modified algorithm for computation of final multiplicity of edge A, B

Data: Vertices $A, B, C; A < B < C$; light position L in homogeneous coordinates

Result: Multiplicity m for one opposite vertex

$X = C - A$;

$Y = (l_x - a_x l_w, l_y - a_y l_w, l_z - a_z l_w)$;

$N = X \times Y$;

$m = \text{sgn}(N \cdot (B - A))$;

Algorithm 4: $\text{CompMultiplicity}(A, B, C, L)$ function used in algorithm 3

After tessellation, we have to transform tessellation coordinates into vertex position of the shadow volume quad in the evaluation shader. The algorithm for its implementation is described in Algorithm 5 and Equations 2.

$$\begin{aligned} A &= (a_x, a_y, a_z, 1)^T \\ B &= (b_x, b_y, b_z, 1)^T \\ C &= (a_x - l_x, a_y - l_y, a_z - l_z, 0)^T \\ D &= (b_x - l_x, b_y - l_y, b_z - l_z, 0)^T \end{aligned} \quad (2)$$

Because caps are not generated, this method can also be used with simpler z-pass algorithm.

2.3 Implementation

All our methods were implemented in Lexolights, an open-source multi-platform program based on OpenSceneGraph and Delta3D, using OpenGL.

Single-pass per-triangle method suffers from inconsistent rasterization of two identical triangles at the same depth but with different winding - depth of fragments from both triangles differs, which resulted in z-fighting artifacts. We had to manually push the front cap's

Data: Vertices **A, B, C, D**, tessellation coordinates $x, y \in \langle 0, 1 \rangle$ and multiplicity m

Result: Vertex **V** in world-space

$P_0 = A$;

$P_1 = B$;

$P_2 = C$;

$P_3 = D$;

$a = \text{round}(x \cdot m)$;

$b = \text{round}(y)$;

$id = a \cdot 2 + b$;

$t = (id \bmod 2) \wedge (\lfloor id/4 \rfloor \bmod 2)$;

$l = \lfloor (id + 2)/4 \rfloor \bmod 2$;

$n = t + l \cdot 2$;

$V = P_n$;

Algorithm 5: This algorithm transforms tessellation coordinates into the vertex of side of shadow volume. Vertices **A, B, C, D** are computed using Equation 2.

fragments into depth of 1.0f, so they would fail the depth test, otherwise we observed self-shadowing artifacts. Bypassing early depth test in rasterization due to assigning depth values in fragment shader causes significant performance loss over two-pass method. This method served as a basis for silhouette-based approach.

For caps generation in silhouette-based method, we used geometry shader and multiplicity calculation, using which we calculated triangle's orientation towards light source via reference edge. It was also necessary for keeping discarding calculations consistent throughout the rendering process of shadow volumes.

Because tessellation factors are limited, at the time of writing, to 64, there is also a limit of maximum multiplicity per edge that this algorithm is able to process. According to equation to calculate tessellation factor $Multiplicity \cdot 2 - 1$, maximum multiplicity of an edge is 32, which should be more than enough for majority of models. But for example well-known Power Plant model (12M triangles) has some edges, which have multiplicity of 128. In that case, they would have to be splitted into more edges.

3 EXPERIMENTS

We compared our methods against already available shadow volumes implementations on modern hardware - robust geometry shader implementation and standard shadow mapping, using which we also tried to evaluate performance against Sintorn's AFSM [Sintorn et al., 2008]. We also tested two-pass per-triangle method against similar geometry shader implementation. For shadow volumes approaches, z-fail was used; shadow map resolution was set to 8k x 8k texels.

Testing platform had following configuration: Intel Xeon E3-1230V3, 3.3 GHz; 16GiB DDR3; GPUs: AMD Radeon R9 280X 3 GiB GDDR5, nVidia

Spheres10x10 Triangles	R280				G680			
	TS		GS		TS		GS	
32400	984	995	490	484	739	825	542	540
67600	921	963	488	487	624	667	494	513
102400	615	729	484	479	491	555	372	402
360000	203	233	270	272	218	228	131	135
1081600	72	88	104	110	82	94	46	49
1440000	56	72	84	91	67	81	36	39
1960000	34	41	59	62	49	58	26	28

Table 1: Performance of two determinism methods measured in FPS on a scene with 10x10 spheres at different triangle count.

GeForce GTX 680 2GiB GDDR5; Windows 7 x64; driver version: 13.12 (AMD), 334.89 (nVidia).

3.1 Testing Scenes

We created a camera fly-through in two testing scenes, each having one point light source.

- Sphere scene: synthetic scene containing adjustable number of spheres (typically 100) with configurable amount of detailness. Fly-through has 16 seconds.
- Crytek Sponza: popular model used to evaluate computer graphics algorithms. 262 267 triangles, 40 seconds.

3.2 Results

Majority of our tests was performed on a sphere scene with adjustable amount of geometry. First, we made a flythrough in a scene containing 100 spheres with different amount of triangles per scene, the results can be seen in Table 1 and graph in Fig. 5.

On GTX680, tessellation using reference edge is the fastest, no matter the number of triangles, although the performance gaps gets smaller with increasing number of triangles in scene. R9 280X showed different results, tessellation was more than 2x faster when the scene contained only 32K triangles but at approximately 300K, geometry shader method took lead.

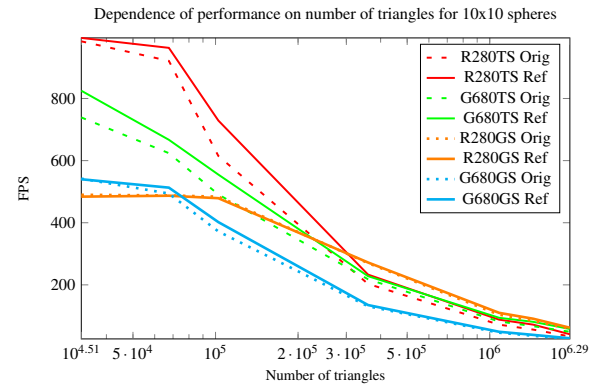


Figure 5: Dependence of performance (FPS) on number of triangles on a scene with 10x10 spheres using original and new deterministic method.

Spheres 1M	R280				G680			
	TS		GS		TS		GS	
Objects								
1	73	92	111	120	106	134	55	60
4	74	94	113	121	101	126	53	58
25	64	76	97	101	76	88	46	50
64	68	76	90	89	61	66	40	43
100	64	70	84	82	58	62	39	42
240	58	55	70	64	50	49	35	36
399	53	48	61	54	36	36	27	28
625	43	38	53	46	29	27	22	22
851	40	44	46	50	24	25	19	20
1250	35	37	28	31	19	19	16	16
2500	23	19	15.1	15.4	12	11	10.8	10.1
3116	21.2	21.5	12.8	12.5	11.1	11.2	9.1	9.2
3920	15.7	14	10.1	10.12	9.1	8.7	7.7	7.5
5100	14.8	14.2	7.8	7.75	8.2	8.2	6.7	6.8
15600	7.45	6.45	3.07	3.14	10.5	9.1	3.6	3.6

Table 2: Dependence on number of objects for spheres scene with 1M triangles. Bold values represent the fastest algorithm/implementation for respective number of objects, per GPU.

We further extended this test to performance dependency on number of objects in a scene while maintaining constant amount of geometry. This measurement was carried out on Sphere scene, having 1 million triangles (with deviation max 2%) in every case. No hardware instancing was used, every object was drawn via separate draw call. Results can be seen in Table 2 and graph in Figure 6.

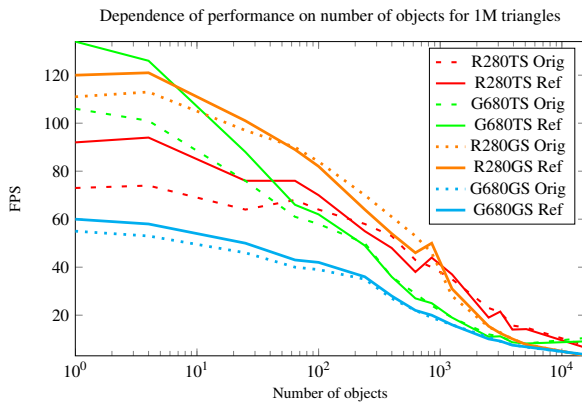


Figure 6: Dependence on number of objects for spheres scene with 1M triangles.

Contrary to previous measurements, tessellation was faster on R9 280X, starting from 10^3 objects, although reference edge was faster only in 40% cases. Moreover, as can be seen in Fig. 6, there is a slight increase in FPS in both geometry shader and tessellation implementations at about 1000 objects on Radeon. On GTX680, tessellation method was faster in every case; eference edge provided increased performance only in a half of measurements, but in all other cases the difference was only 1-3 FPS.

Sintorn in his AFSM paper Sintorn et al. [2008] stated that his per-pixel precise shadow maps are 3-times slower than standard 8Kx8K shadow mapping. In order to evaluate our algorithm against AFSM, we

Spheres10x10	R280		G680	
	TS	SM	TS	SM
Triangles				
32400	995	252	825	245
67600	963	250	667	237
102400	729	244	555	225
360000	233	219	228	190
1081600	88	168	94	135
1440000	72	155	81	115
1960000	41	120	58	103

Table 3: Shadow Mapping vs Tessellation Silhouettes, 10x10 sphere scene, FPS

conducted a measurement against shadow mapping having resolution mentioned above, results of which are in table 3 and graph 7.

Dependence of performance on number of triangles for Shadow Mapping and Tessellation Silhouettes

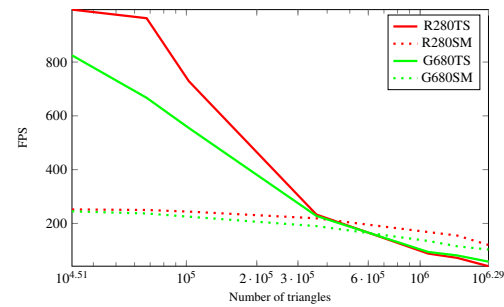


Figure 7: Shadow Mapping vs Tessellation Silhouettes on a scene with 10x10 spheres, measured in frames per second.

Not only we managed to outperform shadow mapping with triangle count up to ~400K triangles, but at almost 2M triangles our method was on par or faster than AFSM - R9 280X dropped to 34% of SM performance whereas GTX680 was only 44% slower than 8K shadow mapping.

We also compared silhouette methods with two-pass per-triangle tessellation implementation and 8K shadow mapping (only on sphere scene, our framework does not support omnidirectional shadow mapping) on Crytek Sponza scene, results in table 4 and graph 8.

One can observe that per triangle tessellation method is even faster than than both geometry shader methods running on Sponza scene. It is also worth noting that per-triangle geometry-shader-based method provides more performance on this scene than silhouette-based approach. On GTX680, the difference between silhouette and per-triangle tessellation method is 122%, whereas on R9 280X card it is only faster by 27%.

With increased amount of geometry in our synthetic test scene, the situation turns around in favor to silhouette methods. Also performance difference between shadow mapping and tessellation on Radeon drops under 1/3 ratio, but GeForce is still able to maintain 43% of SM performance.

4 CONCLUSIONS

We have developed new methods for computing shadow volume silhouettes using tessellation shaders.

Method	R280		G680	
	Spheres	Sponza	Spheres	Sponza
TS Triangle	5.8	102	7.9	83
TS Silhouette	23.7	130	32	185
GS Triangle	3.1	51	4.9	73
GS Silhouette	34	49	14.8	62
SM	93	0	74	0

Table 4: Overall comparison of GS, TS methods and classic 8K shadow mapping on testing scenes - Sponza, and Spheres with 4M triangles. Shadow mapping was not evaluated on Sponza scene (zeros).

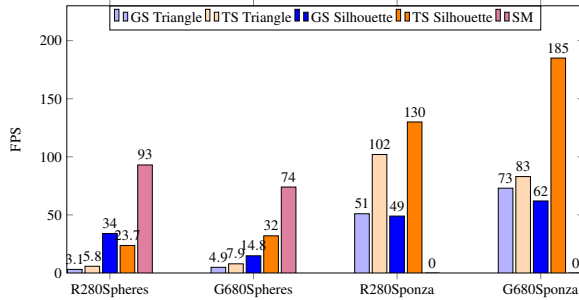


Figure 8: Overall comparison of methods on testing scenes - Sponza and Spheres with 4M triangles.

Our two-pass per-triangle tessellation method is, in some cases, faster than silhouette algorithm implemented in geometry shader, but loses performance as geometry amount in the scene grows. Compared to geometry shader per-triangle implementation, it was faster in every measurement.

The silhouette method is more efficient, and as we have proven in our measurements, mostly in scenes with higher amount of geometry. GeForce GTX680 benefited mostly from this algorithm, being faster than geometry shader silhouette method. As for Radeon R9 280X, geometry shader method is more suitable. Tessellation method on Radeon proved to be faster in Sponza scene, but our synthetic tests on sphere scene showed that it's performance is dominant only up to ~300K of triangles when having multiple objects in the scene, or only up to 15K triangles when only a single detailed object was drawn. In less detailed scenes it was able to outperform nVidia-based card, but only up to aforementioned 300K triangles.

Our robust algorithm was sped up by using a novel method of multiplicity computation, which was able to provide up to 31% performance gain in tessellation method (13.5% in average), maximum speedup in geometry shader was 10.7% with average of 3.4%.

In comparison to standard SM and Sintorn's Alias-Free Shadow Maps (AFSM), our tessellation method provides better performance than 8K shadow maps up to ~400K triangles and then fall to 43% performance of shadow mapping at 4M triangles on GeForce, 34% on Radeon, which is on par or better than AFSM (it's 3-times slower than 8K SM) and is also simpler to implement.

In the future, we would like to see an arbitrary \pm stencil operation in hardware, configurable in shaders, which would allow us to increase the speed of our method even more, due to a lower number of triangles being drawn. We also want to evaluate more hardware platforms and explore GPGPU potential in the field of shadow volumes calculation.

ACKNOWLEDGEMENTS

The work has been made possible thanks to the co-funding by the IT4Innovations Centre of Excellence, Ministry of Education, Youth and Sports, Czech Republic, MŠMT, ED1.1.00/02.0070, V3C - Visual Computing Competence Center, Technology Agency of the Czech Republic, TAČR, TE01020415V3C, and RODOS - Transport systems development centre, Technology Agency of the Czech Republic, TAČR, TE01020155.

REFERENCES

- Brabec, S. and Seidel, H.-P. (2003). Shadow volumes on programmable graphics hardware. *Computer Graphics Forum (Eurographics)*, 2003:433–440.
- Crow, F. C. (1977). Shadow algorithms for computer graphics. In *Proceedings of the 4th annual conference on Computer graphics and interactive techniques, SIGGRAPH '77*, pages 242–248, New York, NY, USA. ACM.
- Donnelly, W. and Lauritzen, A. (2006). Variance shadow maps. In *Proceedings of the 2006 Symposium on Interactive 3D Graphics and Games, I3D '06*, pages 161–165. ACM.
- Everitt, C. and Kilgard, M. J. (2002). Practical and robust stenciled shadow volumes for hardware-accelerated rendering.
- Heidmann, T. (1991). Real shadow real time. pages 28–31. IRIS Universe.
- Kim, B., Kim, K., and Turk, G. (2008). A shadow-volume algorithm for opaque and transparent nonmanifold casters. *J. Graphics Tools*, 13(3):1–14.
- McGuire, M., Hughes, J. F., Egan, K., Kilgard, M., and Everitt, C. (2003). Fast, practical and robust shadows. Technical report, NVIDIA Corporation, Austin, TX.
- Pečiva, J., Starka, T., Milet, T., Kobrtek, J., and Zemčík, P. (2013). Robust silhouette shadow volumes on contemporary hardware. In *Conference Proceedings of GraphiCon'2013*, pages 56–59. GraphiCon Scientific Society.
- Sintorn, E., Eisemann, E., and Assarsson, U. (2008). Sample based visibility for soft shadows using alias-free shadow maps. *Computer Graphics Forum (Proceedings of the Eurographics Symposium on Rendering 2008)*, 27(4):1285–1292.
- Sintorn, E., Olsson, O., and Assarsson, U. (2011). An efficient alias-free shadow algorithm for opaque and transparent objects using per-triangle shadow volumes. In *ACM SIGGRAPH Asia 2011*, SIGGRAPH Asia 2011.
- Stich, M., Wächter, C., and Keller, A. (2007). Efficient and robust shadow volumes using hierarchical occlusion culling and geometry shaders. In Nguyen, H., editor, *GPU Gems 3*, pages 239–256. Addison Wesley Professional.
- Williams, L. (1978). Casting curved shadows on curved surfaces. *SIGGRAPH Comput. Graph.*, 12(3):270–274.

A method of micro facial expression recognition based on dense facial motion data

Yasuhiro AKAGI
Kagoshima University
Korimoto, 1-21-24
Kagoshima
890-8580, JAPAN
akagi@ibe.kagoshima-
u.ac.jp

Hiroshi KAWASAKI
Kagoshima University
Korimoto, 1-21-24
Kagoshima
890-8580, JAPAN
kawasaki@ibe.kagoshima-
u.ac.jp

ABSTRACT

In this paper, we propose a method for recognizing a micro expression which is a small motion appearing on a face by using a high density and high frame-rate 3D reconstruction method. Some studies report that the micro expressions are caused by the change of mental state. If we can recognize the micro expressions, this information could be useful for machines to understand the mental state of a human. With advancements of 3D reconstruction methods, methods have been proposed to reconstruct dynamic objects such as motions of a human's body in high accuracy with high frame rate. Based on the data obtained from the high quality shape reconstruction method, the proposed method recognizes the micro expressions. To detect a part of the face where the micro-expressions are appeared, we propose an experimental estimation of the part. We also report a recognition rate of the change of the mental state using the experimental system.

Keywords

Micro expression, Facial expression, AAM, Facial motion

1 INTRODUCTION

The human face displays much information of an internal (mental) state of a human. By using facial images, many studies have been proposed to estimate the human mental state based on a machine learning. In these facial expression recognition methods, they define 5-8 types of human expressions and detect these expressions at more than 90% recognition rate. This confirms in terms of study that human expressions have certain universality in which connects human emotions.

On the other hand, humans can purposefully control facial expressions different from their emotions. However, when experiments are performed to evaluate a facial expression recognition, whether or not it matches the recognized facial expression and the subject's true emotion at this time is not distinguished, since it is excluded from this recognition problem. Ekman et. al. focuses on a facial expression when the mismatch occurs between the expression and emotion[1]. According to

this work, they point out that when a person shows facial expressions different from their true intentions, there exists a part with small expression change and difference in time to form the expressions. They call the small change as "micro expression". In this study, we verify whether or not a person's change in mental state appears on a face and to figure out the part it is expressed on by focusing on small movements (micro expressions) on the face. To detect the micro expressions, a method which can capture facial motions with high accuracy and high frame rate are used. Then we capture the facial motions by using Kinect[5] (common accuracy and frame rate) to detect the micro expressions. By comparing the recognition results of the micro expressions between the two capturing method, we also show the required accuracy and frame rate to recognize the micro expressions.

2 RELATED WORKS

2.1 Facial expression recognition

There are many studies of the human understanding based on facial expressions, since the human face displays various information including different emotions. Among these, numerous experiments have been performed regarding human expression detection. As methods to detect expressions, using points that display the outline of parts such as a mouth (facial feature points), as well as dividing the face image into small

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

regions and using its uniqueness of the region with luminance gradient have been proposed. For method using facial feature points, there is a method called Active Appearance Model (AAM) that places feature points on the outlines of major facial parts such as eyes and nose, and using their changes as a feature[6]. Ekman et. al. first defined parts called Action Units (AUs) that serve as a index to capture facial movements, then extracted the movements from the image to detect expressions[7]. Kotsia et. al. studied 8 types of AU movements using Support Vector Machine (SVM) to reach over 95% expression detection rate[8].

On the other hand, for methods using uniqueness in small region, Shan et. al. extracted unique Local Binary Pattern as feature vector to be studied with SVM to detect expressions[9]. As a result, they were able to reach 91% identification rate for 7 different types of expressions. According to a survey article regarding expression detection, many methods have been proposed that successfully reached over 90% recognition rate[10], and general expression detection has been in practice. On another note, for the purpose of aiding speech detection from a voice, other methods have been proposed to detect speech from the shape of the mouth[11]. Many experiments have been performed to recognize the information displayed on the human face, by using facial feature points.

2.2 Researches focus on a micro expression

Ekman et. al. focused on the subtle movements displayed on the face[1]. In their research, they reported that when a person tries to display expression different from their emotions, the small movements appearing on their face (micro expression) becomes different from the expression with matched emotion. For the research that utilizes the change in micro expression, a method has been proposed that measures the amount of time it takes for the newborn child to recognize from the initial movement (200ms from the beginning) to the time of expression change detection to identify their development disorder[12]. Su-Jing et. al. proposed a method that detects micro expression from an image[2, 3]. Here, 5 types of micro expressions were detected from a moving image captured at 60fps which resulted in about 30 – 47% recognition ratio and concluded that it needs improvements. Studies about the micro expression recognition are still stages in-development and many studies researched about the method to detect the feature of the micro expressions[4]. Although mechanical detection of micro expressions is crucial for human understanding, its subtleness makes it difficult to detect.

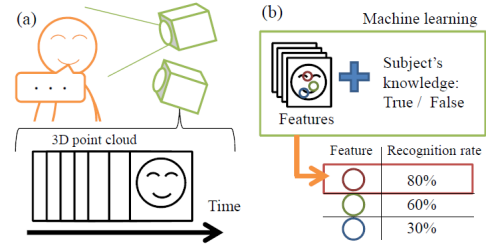


Figure 1: Overview of the proposed system. (a)Measurement of a face during speech. (b)Detection of a part where a micro-expression appears.

2.3 Methods of feature detection of a face

As mentioned in 2.1, in order to detect expressions based on facial feature points, it is necessary to extract high quality feature points from the image. Dollar et. al. achieved error difference of 3 pixel by studying the uniqueness on the image using Random fern[13]. Also, Cao et. al.[14] was able to track feature points in real time based on AAM, from the depth map obtained from devices such as Kinect[5]. To detect the micro expressions, it requires highly precise measurement with high quality face shape data, we propose a method that utilizes the detailed 3D shape data with subtle changes displayed on the face. In our research, by using the highly precise measurement of the 3D shape data, we examine which movement related to human mental state is being displayed on the face. From this result, we experimentally created a system that estimates a simple human mental state, and tested its detection accuracy.

3 OVERVIEW OF THE RECOGNITION TARGET AND ITS PROCESS

Our experiments are performed under the following hypothesis: if the micro expressions are appeared according with the changing of a person 's mental state, by learning the pair of a facial motion and a mental state of a person, a machine learning method can find the relevances between the facial motion and the mental state (fig. 1). This hypothesis agrees with the approach of a facial expression recognition method. In the following sections, we will explain the overview of the proposed method.

3.1 Experimental method and the state of the subject to be detected

In this research, facial motions under different mental states are required to recognize the micro expressions. To collect these motions, we assign the following subjective experiments to capture facial motions (fig. 2). The subject will see one vocabulary every three seconds on the display, and is to read them out loud. After the experiment, the subject is asked whether or not

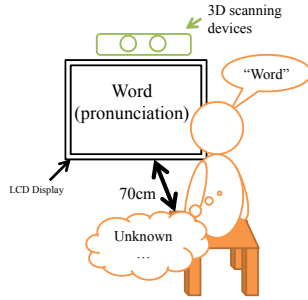


Figure 2: Experiment environment of the task.

they knew the vocabulary asked. Then if it was the first time a subject pronounced a word, even if the subject could assume the meaning of the word from the word, the word is categorized as "unknown". From this experiment, we can record two types of conditions where one knows the vocabulary already and their vocal state is normal; the other is where the subject did not know the vocabulary and speaks under pressure of thought. The reason to choose this method is that the recognition problem becomes simple with only two states (knows/don't know), and the subject can correctly answer the mental state after the experiment. A dataset[15] was used for the vocabulary set.

3.2 Measurement method of 3D shape of a face

In our experiment, we measure the 3D shape of the subject's face from the front while they are answering the questions. For measurement, we use the two types of 3D reconstruction methods; one is that allows high resolution and high frame rate, another is a simple measurement method using Kinect. The facial motion data from the first measurement method is to determine if there are changes in the face when the subject experiences an unknown word, and to detect the region with changes. The second measurement method was done to determine if a simple measurement system can accommodate the proposed detection problem. Table 1 and fig. 3 show the difference between the two measurement methods in respect to measurement preciseness and the difference in the number of measured points. The Table 1 shows the numbers of points when the reconstruction system (camera) is located in the minimum distance from a face. The next section will explain the detection method of the facial area where the micro-expression is appeared from an input point cloud.

3.3 Estimation of the site of micro expression using precise 3D facial motions

According to the research of Ekman et. al. mentioned above[1], the human face has parts that are easy to actively control and parts that are not so easy. It

Table 1: Comparisons of specifications of each measurement method.

	Measurement method	
	High accuracy	Kinect
Resolution(mm)	0.3	2
No. of points	30,000	5,000
Frame rate	200	20

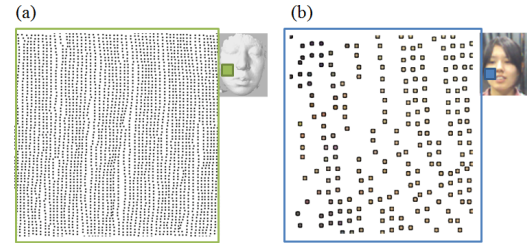


Figure 3: Comparison of point clouds captured by both measurement methods. (a)Points scanned by the high accuracy method. (b)Points scanned by Kinect.

is reported as one example, the characteristic region around the mouth is easily controlled, spaces between eye brows are difficult to control, where the true emotion are often expressed. Since facial movement is likely to have different uniqueness between different parts, the face is divided into mouth region, nose region, and the eye brow region and determined which region best contains uniqueness regarding detection of words used in this experiment. For each region, points are selected randomly from the starting frame of the face with speech, and variation in the depth direction at those positions is recorded. Then, using the feature vector for this variation and by placing supervised learning labels as the unknown words, the difference in the recognition rate is compared (fig. 1(b)). From the result, in the parts with high identification rate, the movement related to the unknown word is considered to be present. The detailed method is stated in section 4.

3.4 Recognition of unknown word with Kinect

As method explained in section 3.3, by using a large number of feature points with high-precision sensor, it is possible to measure the small changes in the face and although it is advantageous for the recognition of subtle movements of the face, it requires special equipment for capturing and also too costly. On the other hand, its measurement accuracy is low and when using sensor with low frame rate, it is possible to not be able to find the subtle movement changes. For this problem, since it is possible to limit the part with unique deformation as mentioned in section 3.3, the change in the recognition rate is examined for the narrow measurement range. Cheaper 3D measurement device, Kinect is

used for this experiment. Ideas of measurement method using the Kinect are described in section 5.

4 STUDY OF SPEECH FEATURE BASED ON HIGH-PRECISION 3D SHAPE DATA

This section experimentally detects which site appears to uniquely show the difference in the shape by recognition of the words during pronunciation using the 3D shape data of the face taken on high speed and high precision.

4.1 Measurement of high-precision facial movement data

In the study presented in this section, it is necessary to extract the movement unaffected by the difference of the words spoken by the subject, and only the movements originating from movements from unknown words. Furthermore, since there is a possibility that a micro expression might have a 'habit' specific to the subject, two patterns of study are performed for the case of extracting individual features for each subject, and for the case of extracting regardless of the subject. For this study, multiple speakers are asked to read the words multiple times, and the face shapes are measured individually. The measurement was performed continuously until 14 known and 14 unknown words appear to be used to be further evaluated. Whether the words were known or unknown is reported from the subject after the measurement (after speech).

4.2 Study of movement data and the data used for the comparison survey

Since there are facial parts that are easy to control and parts that are not so easy[1] we divide a face into three regions: mouth, nose, and eyes for defining the feature vectors of facial movements. This study was done using the shape deformation data in the region of radius 2.5cm from the center position, for the three parts: mouth, nose, and between the eyebrows (fig. 4). We given the center positions of each region manually and 200 points are selected randomly in each region. We define the feature of the facial movement by using the depth directional changes of 100 frames (0.5 sec) at each point. By applying the facial motion tracking method[16], we track the motion of each points. The following measurement was performed for the 14 known and unknown words respectively, with each dataset being represented by 100 reference points. We define the feature vector of each point as Z-axis movements of the point in 100 frames (100 dimension vectors). Also, each feature vector was given a label whether or not the words were known or unknown. Based on these feature vectors and labels, even when

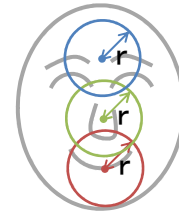


Figure 4: Comparison areas of a motion feature.

given with multi-dimensional data with numerous instructing data, mechanical study is performed that can efficiently learn using two class classifying Local Deep Kernel Learning (LDKL)[17].

4.3 Investigation of the recognition rate when considering individuality of the micro expression

We collect the motion data of 28 words (14 known, 14 unknown) from 6 subjects. The 24 words (12 known, 12 unknown) were used for machine learning and 4 words were used for evaluation. Table 2 shows the recognition ratio for 6 individuals. In any subject's case, the result from the motion features of the area around the nose had the highest detection rate. In particular, 75% or more detection rate was shown for the knowledge of words, and is highly likely that the nose area displays a feature movement. On the other hand, subject B's recognition ratio of unknown words was 46%, which shows that movement feature did not occur. Further, the recognition rate for the mouth region indicates 54 – 65%, but it cannot be said so when it is considered to have two type classification, thus motion resulting from the unknown words were found at the mouth region. This is also true for the region between eye brows.

4.4 Investigation of the recognition rate when not considering individuality of the micro expression

Next, we investigate the recognition ratio using the data without distinguishing characteristic movements for all subjects. Evaluation was done using the set of vocabulary not used in the learning process for each individual. The result is shown on Table 3.

Although the identification rate of the result has increased here, when it is compared with the data used for the learning process in section 4.3, it has decreased. This result shows that micro expression has some levels of individual differences. This is connected to the research mentioned by Ekman et. al. that ability to hide emotions such as a startle is different from subject to subject[1]. Therefore, when trying to make an identification device that uses multiple subjects, a solution is needed to overcome individual differences.

Table 2: Recognition rate of unknown words based on the motion data of the same person (%) . T:Known, F:Unknown words.

	Subject A		Subject B		Subject C		Subject D		Subject E		Subject F		Total	
	T	F	T	F	T	F	T	F	T	F	T	F	T	F
Mouth	62	65	54	65	66	45	77	45	70	13	29	62	60	49
Nose	79	77	77	46	75	78	70	69	78	57	73	74	75	67
Brow	44	5	72	34	41	19	55	18	85	51	71	21	61	34

Table 3: Recognition rate of unknown words based on the all motion data (%) . T:Known, F:Unknown words.

	Subject A		Subject B		Subject C		Subject D		Subject E		Subject F		Total	
	T	F	T	F	T	F	T	F	T	F	T	F	T	F
Mouth	45	54	69	62	52	75	36	54	53	70	24	68	46	64
Nose	72	45	62	73	73	79	75	76	87	31	73	86	74	65
Brow	58	4	50	61	53	15	45	21	84	76	43	30	55	35

5 DETECTION WITH A SIMPLE SYSTEM USING KINECT

From the result of section 4, when identifying the nose area, information to help identify the cognitive word can be obtained. Based on this knowledge, the change of detection rate using Kinect, which has more noise than the measurement method explained in section 4, is researched.

5.1 Measurement data and pretreatment

In addition to the target measurement of the 3D shape, 64 points of feature points can be obtained from the Kinect as the facial feature points[5]. Out of them, by using the 12 points that make up the outline of the nose, it is used as the detection and the study of the depth direction of the point as its feature vector. However, since there is a possibility that measurement errors and noise are contained in the Kinect, by pre-analyzing the motion information for frequency and also removing the signal component of 10Hz or more, variation of the time series have been smoothed out. In order to gather data from all of the seven subjects involved, facial data was measured for the full 60-words vocalization dataset. The known and unknown word ratio is 215:145. The ratio of the known words is structured to be slightly higher.

5.2 Recognition based on motion characteristics using Kinect

As similar to the experiment done in section 4.3, the study and its recognition is done using the same subject. From the 60-word worth of vocalization data gathered from one subject, 12 points that make up the nose and 40 frames (2.0sec) were obtained. Also, 45 words during the study process, and 15 words for evaluation were used. The result is as shown on Table 4.

From this result, although it can recognize the information regarding the unknown words from a partial subject, there are some subjects that couldn't capture any feature points. As stated in section 4.4, for subjects that shows little micro expressions, lowering the measurement accuracy makes it difficult to capture the feature points.

Table 4 also shows the results obtained when using only motion feature vectors for the mouth and eye brow area. As with the identification result based on accurate data, as compared to the nose area, recognition rate has decreased. Within the result, there are some results from the mouth data from subject A and C, subject B's eye brow data that has more than 80% accuracy. Since this is a two class problem, if the detection device tends to put the input data more in one class, one class will end up having higher recognition rate than the other. It is likely that the similar phenomenon occurred in our experiment, since the other ratio was at 43%.

6 CONCLUSIONS

In this paper, by using the dense, accurate facial shape measurement, we propose a method to find subtle yet specific moving parts called micro expression. In order to verify the set of methods to identify the site of micro expression and its accuracy, as a simple recognition problem, vocalizing unknown words was assigned to be further analyzed mechanically. As a result, for the task given to the subject, by using highly accurate data of the facial movement, we were able to identify feature parts that matches with the human mental state, and able to estimate at about 75% accuracy. On the other hand, if the accuracy of the face shape measurement was decreased by few mm and lower the frame rate to 20fps, recognition of the micro expression became troublesome. Further research will improve the experimental method to create micro expression, as well as

Table 4: Recognition rate (%) based on feature vectors around the mouth, nose and brow. T:Known, F:Unknown words.

		Mouth			Brow			Brow		
		T	F	Total	T	F	Total	T	F	Total
S u b j e c t	A	87	43	67	64	78	70	62	0	33
	B	67	33	53	56	60	57	88	14	53
	C	80	40	53	100	50	80	60	50	53
	D	36	55	44	67	67	67	46	42	44
	E	56	50	53	33	67	47	25	29	27
	F	72	0	53	88	29	60	83	22	47
	G	25	73	60	75	54	60	20	70	53

by performing on larger scale subject number, we aim to structure feature movements based on micro expressions.

ACKNOWLEDGMENT

This work was supported in part by NEXT program No.LR030 and KAKENHI No.25870570 in Japan.

7 REFERENCES

- [1] Ekman, P.: Facial expression and emotion. *American Psychologist* **48** (1993) 384–392
- [2] Wang, S., Chen, H.L., Yan, W.J., Chen, Y.H., Fu, X.: Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine. *Neural Processing Letters* **39** (2014) 25–43
- [3] Yan, W.J., Wu, Q., Liu, Y.J., Wang, S.J., Fu, X.: Casm database: A dataset of spontaneous micro-expressions collected from neutralized faces. In: *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on.* (2013) 1–7
- [4] Polikovskiy, S., Kameda, Y., Ohta, Y.: Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor. In: *Crime Detection and Prevention (ICDP 2009), 3rd International Conference on.* (2009) 1–6
- [5] Zhang, Z.: Microsoft kinect sensor and its effect. *MultiMedia, IEEE* **19** (2012) 4–10
- [6] Cootes, T., Edwards, G., Taylor, C.: Active appearance models. In: *Proceedings of the 5th European Conference on Computer Vision, ECCV 98. Volume 1407 of Lecture Notes in Computer Science., Springer Berlin Heidelberg* (1998) 484–498
- [7] Ekman, P., Friesen, W.V.: Facial action coding system. A Technique for the Measurement of Facial Movement. (1978)
- [8] Kotsia, I., Pitas, I.: Facial expression recognition in image sequences using geometric deformation features and support vector machines. *Trans. Image Proc.* **16** (2007) 172–187
- [9] Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing* **27** (2009) 803 – 816
- [10] Bettadapura, V.: Face expression recognition and analysis: The state of the art. *CoRR abs/1203.6722* (2012)
- [11] KOMAI, Y., MIYAMOTO, C., TAKIGUCHI, T., ARIKI, Y.: Phoneme analysis of image feature in utterance recognition using. *MIRU 2010 IS3* (2010) 1771–1778
- [12] ICHIKAWA, H., YAMAGUCHI, M.K.: Dynamic subtle facial expression can be recognized by 6- to 7-month-old infants. *Japanese Psychological Research* **56** (2014) 15–23
- [13] Dollár, P., Welinder, P., Perona, P.: Cascaded pose regression. In: *CVPR.* (2010)
- [14] Cao, Z., Yin, Q., Tang, X., Sun, J.: Face recognition with learning-based descriptor. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.* (2010) 2707–2714
- [15] Kazuyo, T., Hayamizu, S., Ohta, K.: The etl speech database for speech analysis and recognition research. *First International Conference on Spoken Language Processing* (1990)
- [16] Akagi, Y., Furukawa, R., Sagawa, R., Ogawara, K., Kawasaki, H.: A facial tracking and transfer method with a key point refinement. In: *SIGGRAPH Posters.* (2013) 79
- [17] Jose, C., Goyal, P., Aggrwal, P., Varma, M.: Local deep kernel learning for efficient non-linear svm prediction. In: *Proceedings of the 30th International Conference on Machine Learning (ICML-13). Volume 28.* (2013) 486–494

Human-Computer Interaction Using Robust Gesture Recognition

Matthias Endler

Oleg Lobachev

Michael Guthe

University Bayreuth, 95447 Bayreuth, Germany

ABSTRACT

We present a detector cascade for robust real-time tracking of hand movements on consumer-level hardware. We adapt existing detectors to our setting: Haar, CAMSHIFT, shape detector, skin detector. We use *all* these detectors at once. Our main contributions are: first, utilization of *bootstrapping*: Haar bootstraps itself, then its results are used to bootstrap the other filters; second, the usage of temporal filtering for more robust detection and to remove outliers; third, we adapted the detectors for more robust hand detection. The resulting system produces very robust results in real time. We evaluate both the robustness and the real-time capability.

Keywords

human-computer interaction, gesture recognition, computer vision, hand tracking

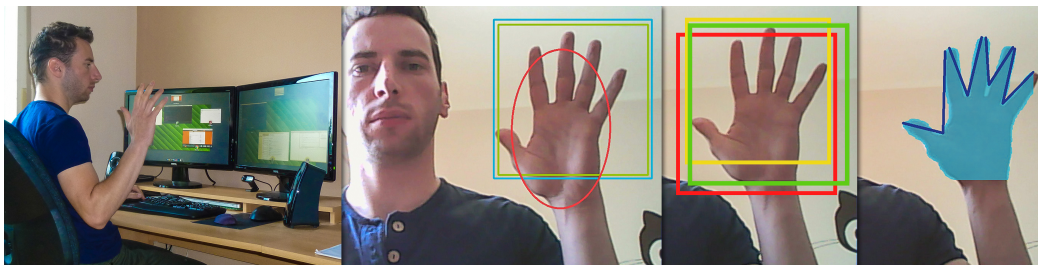


Figure 1: Hand detection and tracking. From left to right: overview of the setup; input frame with Haar-detected bounding box, CAMSHIFT-fitted-ellipse, and bootstrapped-Shape-bounding box; Haar bootstrapping (only once during initialization); final result with hand tracked and fingers counted, used to control applications.

1 INTRODUCTION

Robust, gesture based device control enables a plethora of possibilities to simplify our everyday life, as gestures are an important natural form of human expression. Gestures are a natural part of human interaction. With intrinsic form of communication for human-machine interaction one could intuitively control various devices.

A popular approach in gesture recognition is using special hardware. Gesture recognition was always highly motivated by gaming applications. One of the most prominent examples include Microsoft Kinect that utilizes depth-mapping to augment the vision process. In a contrast to such approaches, we utilize unaugmented live 2D input from a webcam. Our goal is to combine

multiple gesture detection methods in a stable pipeline to achieve robustness and real-time capability.

This work focuses on hand movement. Our system falls in the category of dynamic action recognition – we consider the temporal aspect. However, before we can compute the hand movement, we need to secure that the object we are tracking is indeed user's hand.

We do not apply machine learning techniques, but combine and augment known gesture detection algorithms. Each method receives input frames and returns a detected shape (in some form) and a confidence level. These results are combined to maximize performance. We use the Haar classifier [11, 18], CAMSHIFT [5], Shape [4], and Skin [13, 16]. An overview is shown in Figure 1. To further increase the performance we have modified the algorithms, as detailed below.

Some of these algorithms (we denote *detectors* or *filters* from now on) require prior training. We solve the problem with *bootstrapping*. As soon as the simplest detector produces results with sufficient confidence, these are used to train the more sophisticated detectors. This is an iterative process, that nevertheless completes quite fast and does not interfere with the actual interaction session. The ‘filter cascade’ allows us to execute the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

final training on the actual data, hence no discrepancy arises between the training and interaction data sets.

The contributions of this paper include: An iterative bootstrapping approach to supply the filters with required initial data, basing on the same data stream the detection will subsequently operate; Modifications of the detector algorithms to optimize their performance for hand detection; Evaluation of the resulting software, especially in terms of real-time capability and robustness of the detection process.

2 RELATED WORK

Directly related to our approach is gesture recognition in games, like the Microsoft Kinect [20], however the Kinect uses additional depth information that we do not utilize. Wang et al. [19] compared various classifier combinations for hand gesture recognition. Their input was acquired with two cameras, we use only one webcam; in a further contrast we operate on the live webcam stream and continuously apply some temporal predicates.

We agree with the general state of the art (e.g. [15]) on the segmentation of the hand, however, we use a *combination* of various methods like Viola–Jones classifier [18] (from now on called Haar) and the CAMSHIFT algorithm [5] instead of their phases two and three of [15]. The advances in sign language recognition [1, 3, 9] are less relevant to us, as for these approaches two hand gestures need to be recognized; hand positions relatively to each other and to the face matter. Varying languages are a further difficulty level.

We refer to two surveys [8, 14] of recognition and tracking methods. Quite related to our work are various keypoint and feature detection algorithms [2, 10, 12]. A recent comparison of binary features is [7]. Vezhnevets et al. [17] survey early *skin detection* methods. Similar to many others we also use the HSV color space. Tripathi et al. [16] use YCbCr color model. Our Skin reimplements Pulkit and Atsuo [13].

3 FILTERS

Our system consists of several image filters used as hand detectors. We divide the detectors into position and the feature cascades. Our approach combines multiple detectors: Haar, CAMSHIFT, Shape, and Skin, an overview is in Figure 2. The first cascade is responsible for obtaining the location of a hand in an image, the second exposes hand features in a given area of an image. It requires the data from the first, position cascade to function properly. Even further, the CAMSHIFT and Shape filters from the position cascade are *bootstrapped* using the data from the Haar filter. Here we describe the separate filters, Section 4 discusses their combination.

Our Haar filter was introduced as Viola–Jones cascade-classifier [18], with further extensions (e.g. rotation) [11]. It produces a magnitude of features, similar to Haar

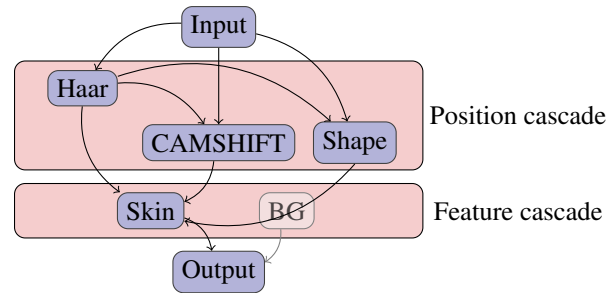


Figure 2: A schematic representation of our approach. First cascade filters provide information to second cascade. Background subtraction (BG) is future work.

bases and then utilizes AdaBoost [6] to select the few meaningful specimen. Haar combines multiple weak classifiers to a stronger one [11]. The *cascade of classifiers* is a chain, where each classifier detects almost all positive features and removes a significant part of negative features – each classifier removes a different kind of negative features, thus the chain yields highly accurate results. Haar is typically used for face recognition, the detection rate for frontal faces can be as high as 98%. While we also used it to eliminate the face in the input image in the pre-processing stage, *our* critical usage of Haar stems on a training set with hand images.

Continuously Adaptive Mean-Shift (CAMSHIFT) [5] basically computes a hue histogram of the tracked area and, in the new frame, *shifts* the tracked area based on probabilities of the pixels to be part of the tracked area. The shift goes to the center of gravity of most probable pixels. Then the histogram is updated. CAMSHIFT requires the initial area to track. We basically use the part of the image for that Haar is confident enough and *track* it not merely with Haar, but also with CAMSHIFT and Shape, see Section 4 for details.

A further (very popular) filter is template matching [4]. Given an base image (the *haystack*) and a smaller template image of size $w \times h$, (the *needle*), we compare patches of size $w \times h$ with the template image, the needle appears at the patch position with the maximum value. Our Shape filter detects the hand position in this manner. We utilize only the red channel and use the normalized cross-correlation metric. The shape of the hand is obtained from the Haar filter.

A single skin color model can be successfully used for most human beings in HSV color space [5]. The hue value won't vary much, only the saturation needs to be adjusted individually. We adjusted saturation and brightness for the varying real-life lightning configurations, our Skin detector utilizes these data.

4 TRACKING SYSTEM

We aim for a real-time capable system that is very robust in terms of detection performance and is easily usable

as a software component in larger applications. The system should work without the need for long calibration steps, it should be platform independent and mostly unobtrusive. After our system has registered the gesture, an action is triggered.

Given an average frame rate of 24 fps, a calculation time of about 40 ms sets a challenging boundary. Even more, as delays would dramatically affect user experience. We optimized the code for real-time usage. The number of false positives needs to be kept to a minimum. Extensibility is also required. We used a modular plugin system for the detection algorithms and a high-level scripting language (Python) with the highly optimized C++ library OpenCV.

All filters were adjusted to improve hand detection. Haar has proven very robust in our tests, we use it to detect the initial hand position. During early tests, we experienced many false-positives with a face. Therefore, a cascade trained on frontal and profile faces was used to remove the face in the image (if any), improving the hand detection rate significantly. We further use two separate hand cascades. One detects open hands and the other one fists. All three cascades run inside one detector. Haar often detects the same object at slightly different positions, called *neighbors*. The more neighbors, the more likely is the correct detection. We dynamically adjust the minimum number of neighbors to retain a candidate.

Haar performs very good on simple, plain-colored backgrounds, but it falls short on complex backgrounds and in difficult lighting conditions. It is quite vulnerable to hand rotations. CAMSHIFT can adapt to changing brightness and hand rotation with backprojection. It provides stable results for complex backgrounds. CAMSHIFT is prone to errors, when other skin-colored objects (e.g. the face) are visible. Shape shows a similar behavior: Our modification made it agnostic to lightning. Like Haar, it won't detect tilted hands. To compensate for varying light conditions, Haar and Shape work on normalized grayscale images; CAMSHIFT and Skin filter operate in HSV.

We assume that the bounding boxes of the hand in two consecutive frames need to overlap. We use a memory buffer for a few previous detected hand positions for each filter and discard outliers. This approach facilitates a *temporal link* between separate frames.

CAMSHIFT and Shape are *bootstrapped* with data from Haar filter. Haar bootstraps itself – we use an initial calibration phase when the hand is not moved. The consecutive bounding boxes, detected by Haar should overlap (see Figure 1). Such bootstrapping functions surprisingly good and is one of the novelties of our approach. Haar uses cascade data, pre-trained for hand detection, but Haar is not training its cascades on the hand of the user! Just holding the hand in front of the webcam for few seconds suffices to produce then confident results

with Haar. The hand region is passed to CAMSHIFT and Shape, so they can perform the tracking. In other words, with our bootstrapping technique we are able to track not *some* hand, but exactly the hand of the current user. This ensures fast and consistent operation.

5 RESULTS

We evaluate both the robustness and the real-time capability of our implementation. We use a MacBook Pro with a 2.4 GHz Intel Core 2 Duo, 4 GB of RAM and its integrated camera, Python 2.7, and OpenCV 2.4.6.

We test our system in various repeatable conditions of different 'difficulty grade'. We recorded our typical gestures in various light conditions, with varying complexity of the background and at various speed of gesturing. The videos are publicly available under <http://bit.ly/R6Owu6>. The background complexity varies between simple background, skin-colored background, complex background, and mirroring. For the tests below we used two gestures: 'Exposé' and 'Move.' Table 1 presents the assessment. We saw a good performance almost everywhere; underexposure and highly specular background were expectantly problematic.

#	Background	Lighting conditions	Gesture	Speed	Result
1	Simple	Normal	Exposé	Slow	+
2	Simple	Normal	Exposé	Fast	+
3	Simple	Overexp.	Move	Slow	+ ¹
4	Simple	Overexp.	Exposé	Slow	+
5	Simple	Underexp.	Exposé	Slow	+
6	Skin-colored	Underexp. (Noise)	Exposé	Fast	+ ²
7	Moving	Changing	Exposé	Slow	+
8	Reflections	Underexp.	Exposé	Fast	+ ²
9	Reflections	Underexp.	Move	Slow	- ³
10	Reflections	Normal	Move	Slow	+ ^{1,2}

Table 1: Results of the robustness evaluation.

The *real-time capability* was achieved. In our visual tests, the system worked fluidly, at significantly faster-than-normal frame rate. We noticed no disturbance when the system was working on live video feed. We used the machine's integrated camera and also experimented with an external webcam. The video was sampled at 24 frames per second.

We have benchmarked a typical video sequence for a drag and drop 'Move' gesture (video #3 from above); the results are shown in Table 2. This table serves as a quantifiable comparison of our method and existing work: how would the filters perform solely? We observe that all filters are highly real-time capable.

Combining the *worst case* timings of all filters, we theoretically achieve at least 18 fps on a quite dated machine

¹ With prior bootstrapping.

² After some tweaking.

³ Haar did not correctly identify the hand.

Filter	Frames per second			
	mean	median	variance	min.
Haar	107.74	109.98	331.46	60.51
CAMSHIFT	394.4	403.9	1402.99	161.5
Shape	125.62	130.11	198.81	36.98
Skin	774.4	732.5	48210.8	157.7

Table 2: Benchmark results. We show the hypothetical frame per second rate for each filter, if executed solely.

using a single threaded implementation. Note that the actual minimum frame rate is even higher (45 fps), so the method is definitively real-time capable.

6 CONCLUSION

We have presented a new combination of several detectors for the robust, real-time hand gesture recognition. We have modified and adapted multiple methods from computer vision: Haar, CAMSHIFT, Shape, and Skin filters. They are typically used for face detection, but we adapted them to hand recognition. We have *composed* them into a working and robust system. Our contribution includes: Chaining the filters into two cascades: one for position detection and one for the shape of the hand; Temporal heuristics and position consensus remove detection outliers, thus improving robustness. Our system has a short-time memory. All filters need to agree on the roughly the same area, a complete outlier is discarded; The filters are bootstrapped, hence we always operate on recent and relevant data.

The *bootstrapping* is an important trait of our system. Most filters require some initial images to track and/or compare with. We iterate one filter until it reaches sufficient confidence levels and then use the successfully detected hand position as input for the further, advanced filters. Such a technique proves to be very stable, as it adapts the whole chain to the particularities of the current user (hand shape, hand size, skin color, etc.) and the current setting (e.g. background, light conditions, white balance). This also increases the robustness.

Our system works with a stock webcam on an inexpensive consumer computer hardware. The average filter performance was well real-time on a quite outdated machine. One important future goal is to port our software to systems with little processing power (such as embedded devices or smartphones). With background subtraction we would be able to reduce the search window and hence to achieve better performance. Threading would greatly improve the real-time performance by executing independent filters in parallel.

A crucial issue is the minimization of false positives. We have already improved this point, but it would be possible to improve even more. On that account, we want to implement more heuristics and algorithms, which help

the system track hands despite the occlusion. Of course, simply detecting more gestures is also important.

7 REFERENCES

- [1] O. Aran, T. Burger, L. Akarun, and A. Caplier. Gestural interfaces for hearing-impaired communication. In D. Tzovaras, editor, *Multimodal User Interfaces*, pages 219–250. 2008.
- [2] H. Bay, T. Tuytelaars, and L. Gool. SURF: Speeded up robust features. In *ECCV '06*, LNCS 3951, pages 404–417. Springer, 2006.
- [3] B. C. Bedregal, A. C. R. Costa, and G. P. Dimuro. Fuzzy rule-based hand gesture recognition. In *Artificial Intelligence in Theory and Practice*, IFIP 217, pages 285–294. Springer US, 2006.
- [4] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE T. Pattern. Anal.*, 24(4):509–522, 2002.
- [5] G. R. Bradski. Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal*, (Q2), 1998.
- [6] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Computational Learning Theory*, LNCS 904, pages 23–37. Springer, 1995.
- [7] J. Heinly, E. Dunn, and J.-M. Frahm. Comparative evaluation of binary features. In *ECCV '12*, LNCS 7573, pages 759–773. Springer, 2012.
- [8] E. Hjeltnäs and B. K. Low. Face detection: A survey. *Comput. Vis. Image. Und.*, 83(3):236–274, 2001.
- [9] A. A. Kindiroglu, H. Yalcin, O. Aran, M. Hruz, P. Campr, L. Akarun, and A. Karpov. Automatic recognition fingerspelling gestures in multiple languages for a communication interface for the disabled. *Pattern Recognit. Image Anal.*, 22(4):527–536, 2012.
- [10] S. Leutenegger, M. Chli, and R. Siegwart. BRISK: Binary robust invariant scalable keypoints. *ICCV '11*, pages 2548–2555. IEEE, 2011.
- [11] R. Lienhart, A. Kuranov, and V. Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. LNCS 2781, pages 297–304. Springer, 2003.
- [12] D. Lowe. Object recognition from local scale-invariant features. volume 2 of *ICCV '99*, pages 1150–1157. IEEE, 1999.
- [13] K. Pulkit and Y. Atsuo. Hand gesture recognition by using logical heuristics. Technical Report 25, Japan Advanced Institute of Science and Technology, School of Information Science, 2012.
- [14] S. S. Rautaray and A. Agrawal. Vision based hand gesture recognition for human computer interaction: a survey. *Artif. Intell. Rev.*, pages 1–54, 2012.
- [15] E. Stergiopoulou and N. Papamarkos. Hand gesture recognition using a neural network shape fitting technique. *Eng. Appl. Artif. Intell.*, 22(8):1141–1158, 2009.
- [16] S. Tripathi, V. Sharma, and S. Sharma. Face detection using combined skin color detector and template matching method. *Int. J. Comput. Appl.*, 26(7):5–8, 2011.
- [17] V. Vezhnevets, V. Sazonov, and A. Andreeva. A survey on pixel-based skin color detection techniques. In *GraphiCon, ICCGV'03*, 2003.
- [18] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *CVPR '01*, pages 1:511–518. IEEE, 2001.
- [19] G.-W. Wang, C. Zhang, and J. Zhuang. An application of classifier combination methods in hand gesture recognition. *Math. Probl. Eng.*, 2012.
- [20] Z. Zhang. Microsoft Kinect sensor and its effect. *IEEE Multi-Media*, 19(2):4–10, 2012.

Particle systems-based riverbed modelling over a terrain with hardness layer

Korneliusz K. WARSZAWSKI
Faculty of Electrical Engineering,
Computer Science and Telecommunications
University of Zielona Gora
ul. Szafrana 2
65-246 Zielona Gora, Poland
k.warszawski@weit.uz.zgora.pl

Slawomir S. NIKIEL
Institute of Control
and Computation Engineering
University of Zielona Gora
ul. Szafrana 2
65-246 Zielona Gora, Poland
s.nikiel@issi.uz.zgora.pl

ABSTRACT

This paper proposes a method that applies particle systems to simulate results of hydrological erosion caused by spout, like riverbeds. The terrain model is divided into two layers. The first one stores heights data (typical height-field) while the second is reserved for hardness data. This data structure enables fast and simple implementation of terrain deformation. We present the construction of a particle system terrain modifier, its main attributes and how they influence the final product of the modelling process. The proposed technique behaves like in classical particle systems. It uses emitter as element that control starting location, direction and quantity of particles in a given simulation environment. We choose parameters for particles such as: the current position, directional angle, linear velocity, rotation angle, rotation velocity and the size which define its zone of influence for landscape modification processes. Each emitted particle is moving (rolling) over the surface of terrain structure making deformations at its current position. Scale of the modifications depends on particle parameters and landscape structure susceptibility for modifications process under the particle influence zone. The proposed method is not intended to simulate physically erosion process, but focuses on its results for exploitation in virtual environments in real-time simulations and rapid prototyping of virtual terrains.

Keywords

particle systems, terrain surface modelling, riverbed generation, hardness-field, virtual environment.

1 INTRODUCTION

The achievement of an ideal form in a virtual landscape is possible as far as developers, artists and virtual world builders spend a lot of time manually deforming polygon-meshes. Alternatively, an acceptable level can be obtained much faster by automated techniques. Applications of those techniques have been used in electronic systems with elements of virtual environment for military and civilian training courses, digital entertainment and game developing [Ric99a, Bon05a, Sme10a].

The foundations of mostly generation techniques of terrain modelling are based on self similarity fractal algorithms. Traditional method uses Madelbrot's Midpoint Displacement algorithm and was proposed by Fournier, Fussell and Carpenter [Fou82a]. Initial height-field grid

has 2x2 resolution and in recursive subdivision, the method increases terrain model precision by calculation of height values of newly generated height-field nodes as the averaged height of the neighbour points displaced by a random offset. The subdivision part of the algorithm faced several modifications. Mandelbrot and Musgrave proposed the Hexagon Subdivision and Miller presented both Diamond-Square and Square-Square subdivisions algorithms. All proposed modifications give alternative models for selection of neighbour points, but the main idea of algorithm still remains unchanged [Mil86a, Mus89a, Koh92a, Sal02a].

Virtual scenery can be simply improved by adding rivers, by simulations of geological structure eroded by flow of spout with use of method based on Computational Fluid Dynamics (CFD). This techniques simulate natural flow of water based on physical model [Fos01a, Cha06a]. Results of simulations are visually acceptable but cost of computations is very high and disqualify this techniques for interactive terrain synthesis.

The real-time method for river-based or rainfall erosion models was presented by Nagashima [Nag97a]. Starting point of the method is put on top of the midpoint

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

displacement algorithm, which sets the base edge of erosion. In the following steps the method makes deeper modifications as erosion process proceeds. Teoh [Teo08a] extends procedural method to simulations of coastal erosion, river meanders and river delta formation.

Particle systems are an alternative approach to chaos-based terrain modelling. The technique was originally developed for computer graphics by Reeves [Ree83a] as a fast method for real-time modelling of object with irregular, dynamically changing objects with no defined surface like clouds, fire or explosion. It can be also used for modelling of vegetation and foliage. Proposition of parallel implementation was done Sims [Sim90a] and was adopted by McAllister [McA00a] and Kolb, Latta and Rezk-Salama [Kol04a] for their descriptions of particle systems. Using particle systems for terrain surface modelling, gives us tools to generate complex height-field data, similar to mountains or island-like forms [War08a]. Benes and Roa [Ben04a] uses the technique to simulate sand (or snow) movement and its interaction with other objects in a virtual environment. As shown by research of Benes [Ben08a] and Kristof, Benes, Krivanek and Stava [Kri09a] particle systems can be successfully adopted in simulation of erosion processes. Due to its characteristics the technique can be used as landscape deformer enriched with canyons or riverbeds.

2 TERRAIN MODEL

We use terrain model which is constructed from two layers. The first layer is a standard height-field. The second layer corresponds to the associated hardness-field [War12a]. The resolution, heights and hardness records on both layers are corresponding mutually. The terrain layers initial data can be derived from a file of real landscape data, e.g., Digital Elevation Map (DEM) or Geographic Information System (GIS). It can be also modelled by any automated method, when this information is unavailable or cost of its extraction is unprofitable and procedural results are quite acceptable.

In the Figure 1 we can see the rendered island based on height-field layer, generated by particle downfall algorithm [War09a]. The model was rendered in Teragen. Figure 2 shows artificially generated hardness layer, with use of the Poisson Faulting hardness synthesis algorithm [War12a]. The sample layer is in 6th class, which means that it contains information about distribution of six types of materials with different hardness value (each material is depicted with different shade in greyscale).

3 PARTICLE SYSTEM

The proposed particle system consists of a collection of particles and an emitter, which controls starting loca-

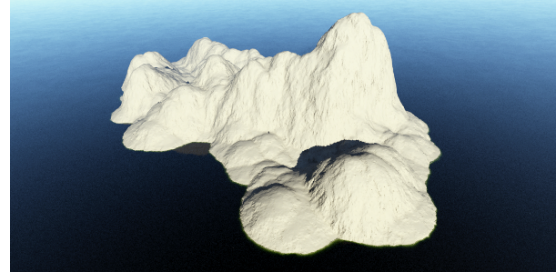


Figure 1: Sample of height-field based island.

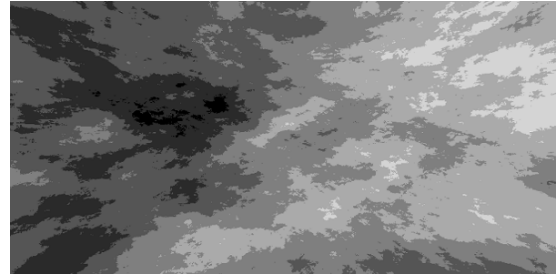


Figure 2: Sample of generated hardness-field.

tion, direction and number of active particles in the system environment. Each particle is described by its position, directional angle, linear velocity, rotational angle, rotational velocity and size.

The particle starting position is selected randomly from area defined by the emitter window. Particle destruction occurs in two cases. Firstly, when its current position exceed bounds of the system workspace. Secondly, as a result of collision with existing terrain fragment. In addition, the collision causes modification (erosion) of the terrain fragment. The erosion zone is related directly to a collided particle size.

4 RIVERBED MODELLING

Setting the emitter position on a side of virtual terrain defines the starting point of the modelled riverbed and its width is defined by the size of the emission window (see Figure 3).

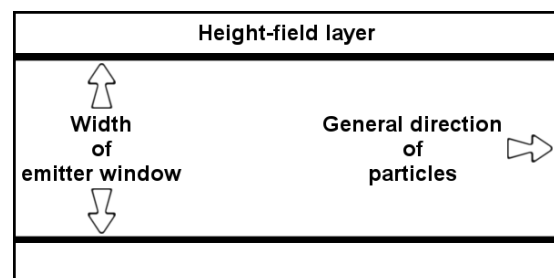


Figure 3: Outline of modelled riverbed.

For our simulation process we supposed that all particles move over the landscape surface. Therefore, we can project their trajectory calculations to a planar equation. Let (x_o, y_o) be a position of point (O) around which given particle (P) rotates. Let (v) be the particle

linear velocity and (α) be its current linear angle. Let (r) corresponds to the particle offset to the (O) point and (β) be its current rotational angle. Then new position for the particle (P) defined as pair of coordinates (x, y) can be acquired with use of Equation 1 and its geometric representation is shown in Figure 4.

$$P(x, y) = \begin{cases} x = x_o + v * \cos(\alpha) + r * \cos(\beta) \\ y = y_o + v * \sin(\alpha) + r * \sin(\beta) \end{cases} \quad (1)$$

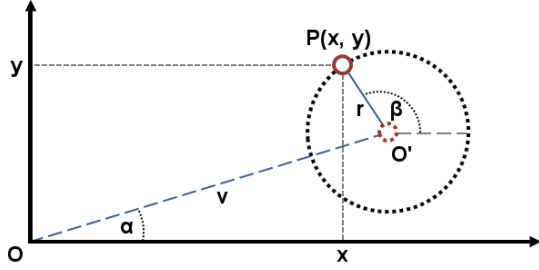


Figure 4: Geometric representation of rotational particle position in \mathbb{R}^2 space.

With the particles movement, terrain is deformed. The strength of these modifications depends on particle parameters and landscape erosion resistance. Let (P) be a particle at position defined as (x, y) and (s) be its size. Let (i, j) be a coordinate at height-field layer. Then the decreasing factor (Δh) for this coordinates can be acquired with Equation 2.

$$\Delta h_{i,j} = \sqrt{s^2 - ((i-x)^2 + (j-y)^2)} \quad (2)$$

The main feature of decreasing factor (Δh) is that it has positive value if the height-field cell is inside the particle zone of erosion and it is negative outside. For our research, we assumed that each height-field cell with positive decreasing factor is a subject to erosion procedure. Let (h) be a height-field layer cell at defined position (i, j) and (Δh) be its decreasing factor. Let (d) be a hardness-field layer cell at the same coordinates. Then the new height value (h') can be acquired with Equation 3.

$$h'_{i,j} = \begin{cases} h_{i,j} - \Delta h_{i,j} * d_{i,j}, & \text{if } \Delta h_{i,j} > 0 \\ h_{i,j}, & \text{otherwise} \end{cases} \quad (3)$$

5 CONCLUSIONS

The flow of particles in a virtual environment is similar to spout. The appropriate simulation of interacting particles for virtual terrain scenery makes it possible to use these techniques to model landscape structures subjected to permanent influence of the hydrological erosion process. The parametric algorithm makes it possible to adapt those results for a satisfactory level of the

modelled terrain. Selection of suitable system attributes enables us to simulate the effects of liquid dynamics on the basic landscape surface.

The performance of the proposed technique depends mostly on the number of particles and decreases while this number increases. Computational complexity was estimated as $F(n) = 6n + 4$, where n is a number of particles. This property of the algorithm enables landscape forming in near real-time simulation, thus it offers interactive terrain sculpting.

In Figure 5 we can observe comparison of results of simulation. Top sub-figure shows riverbed achieved without particle rotation. The other two sub-figures presents model where rotational particles were applied (with and without water layer). All simulations were performed with identical pre-generated 5th class hardness-field layer. In Figure 6 we presents other models generated with the method. All samples were rendered in Terragen.

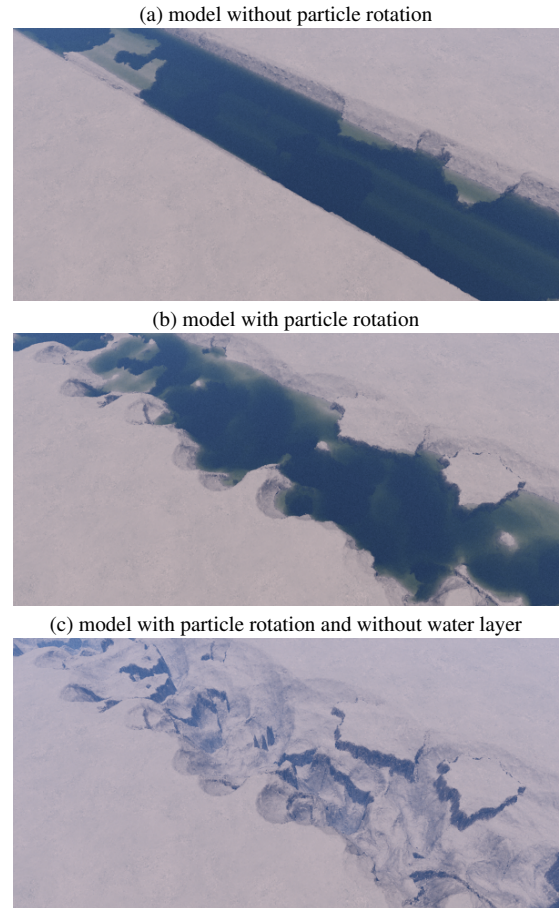


Figure 5: Sample of riverbeds generated over plain height-field with pre-generated hardness records.

Further work will focus on the definition of a fully deterministic method, which could be used to determine flood routes or for optimal regulation of riverbeds. Next, we plan extend this technique to terrain models based on voxel representation.

6 REFERENCES

- [Ben04a] Benes, B., and Roa, T. Simulating Desert Scenery. In Conf.proc of the 12th International Conferences in Central Europe on Computer Graphics, Visualization and Computer Vision, pp. 17-22, Plzen-Bory, Czech Rep., 2004.
- [Ben08a] Benes, B. Real-Time Erosion Using Shallow Water Simulation. In Conf.proc of the 4th Workshop in Virtual Reality Interactions and Physical Simulation, pp. 43-50, Dublin, Ireland, 2007.
- [Bon05a] Bonk, C.J., and Dennen, V.P. Massive multi-player online gaming: A research framework for military training and education. Technical Report for Office of the Under Secretary of Defense for Personnel and Readiness, 2005.
- [Cha06a] Chae, D-J, and Park, J-H. An active model of water movement by activity-based method for simulation of the virtual environment. International Journal of Information Technology, Vol. 12, No. 5, pp. 80-87, 2006.
- [Fos01a] Foster, N., and Fedkiw, R. Practical animation of liquids. In Conf.proc of the 28th annual Conference on Computer Graphics and Interactive Techniques, pp. 23-30, Los Angeles, USA, 2001.
- [Fou82a] Fournier, A., Fussell, D., and Carpenter, L. Computer rendering of stochastic models. Communications of the ACM, 6 (25), pp. 371-384, 1982.
- [Koh92a] Koh, E-K, and D.D. Hearn. Fast Generation and Surface Structuring Methods for Terrain and Other Natural Phenomena. Computer Graphics Forum, Vol. 11, No. 3, pp. 169-180, 1992.
- [Kol04a] Kolb, A., Latta, L., and Rezk-Salama, C. Hardware-based simulation and collision detection for large particle systems. In Conf.proc of the ACM SIGGRAPH/EUROGRAPHICS Conference on Graphics Hardware, pp. 123-131, Grenoble, France, 2004.
- [Kri09a] Kristof, P., Benes, B., Krivanek, J., and Stava, O., Hydraulic Erosion Using Smoothed Particle Hydrodynamics. Computer Graphics Forum, Vol. 28, No. 2, pp. 219-228, 2009.
- [McA00a] McAllister, D.K. The design of an API for particle systems. Technical Report, Department of Computer Science, University of North Carolina at Chapel Hill, 2000.
- [Mil86a] Miller, G.S.P. The definition and rendering of terrain maps. Computer Graphics, Vol. 20, No. 4, pp. 39-48, 1986.
- [Mus89a] Musgrave, F.K., Kolb, C.E., and Mace R.S. The synthesis and rendering of eroded fractal terrains. Computer Graphics, Vol.23, No.3, pp.41-50, 1989.
- [Nag97a] Nagashima, K. Computer generation of eroded valley and mountain terrains. The Visual Computer, Vol. 13, No. 9-10, pp. 456-464, 1997.
- [Ree83a] Reeves W.T. Particle Systems-A technique for modelling a class of fuzzy objects. Computer Graphics, Vol.17, No.3, pp. 359-375, 1983.
- [Ric99a] Rickel, J., and Johnson, W.L. Virtual humans for team training in Virtual Reality, in Conf.proc. World Conference on AI in Education, pp.578-585, Le Mans, France, 1999.
- [Sal02a] Sala, N., Metzeltin, S., and Sala, M. Applications of mathematics in the real world: Territory and landscape, In Conf.proc of the International Conference the Humanistic Renaissance in Mathematics Education, pp. 326-333, Terrasini, Italy, 2002.
- [Sim90a] Sims, K. Particle animation and rendering using data parallel computation. In Conf.proc of the 17th annual Conference on Computer Graphics and Interactive Techniques, pp. 405-413, Dallas, USA, 1990.
- [Sme10a] Smelik, R., Tutenel, T., De Kraker, K.J., and Bidarra, R. Declarative terrain modeling for military training games. International Journal of Computer Games Technology, pp. 2:1-11, 2010.
- [Teo08a] Teoh, S.T. River and coastal action in automatic terrain generation. In Conf.proc of the International Conference on Computer Graphics and Virtual Reality, pp. 3-9, Las Vegas, USA, 2008.
- [War08a] Warszawski K., and Zawadzki, T. Building a particle system for virtual landscape generation, in Conf.proc. International Interdisciplinary Technical Conference of Young Scientists-InterTech, pp.340-343, Poznan, Poland, 2008.
- [War09a] Warszawski, K. Ground from smoke: Using Particle Systems for terrain modeling in C#. Game Developer Magazine, Vol.16, No.3, pp.15-21, 2009.
- [War12a] Warszawski, K., Nikiel, S., and Zawadzki, T. Hardness data synthesis for height-field based landscape models. In Conf.proc of the 18th Surface models for geosciences, pp. 275-284, Ostrava, Czech Rep., 2012.

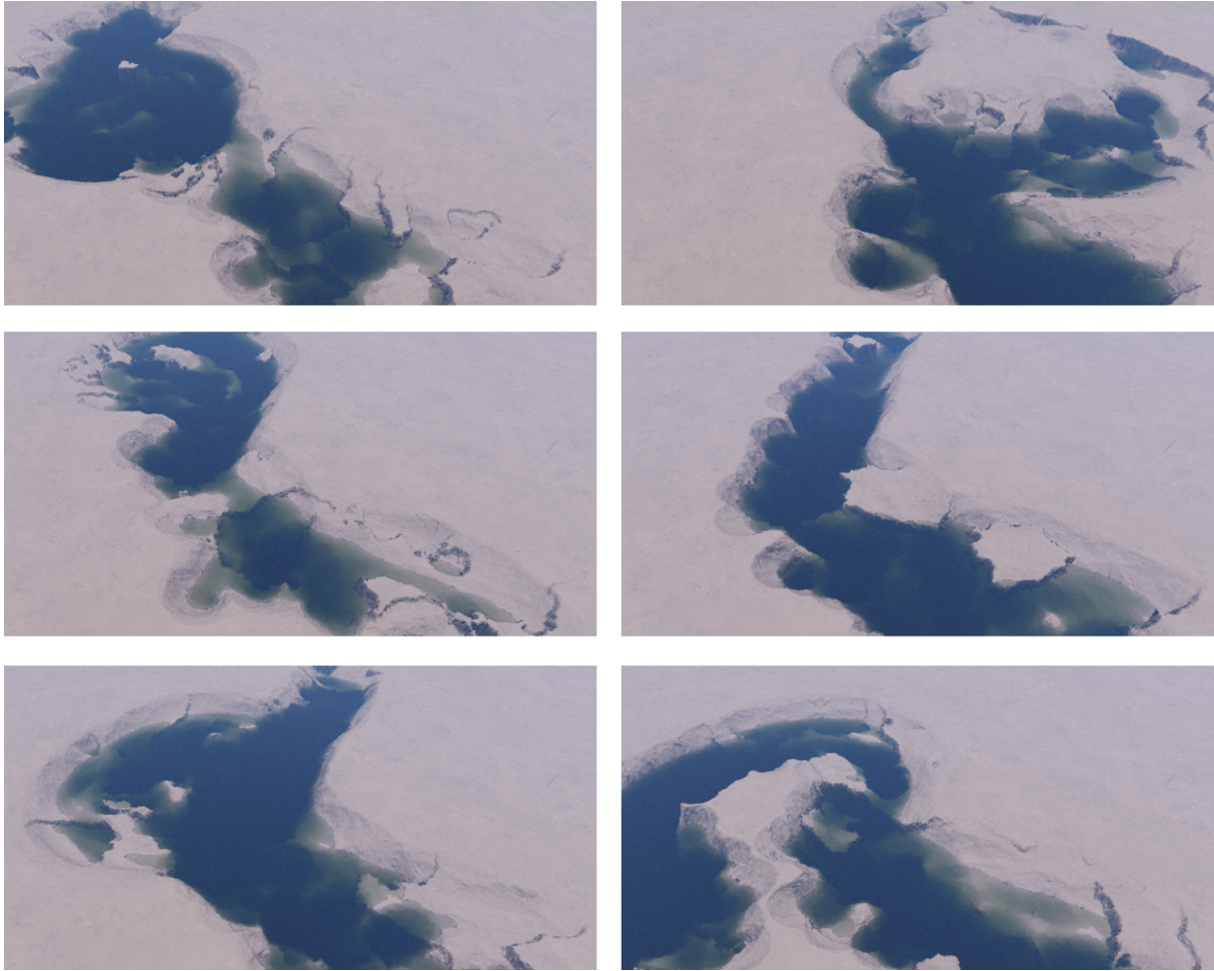


Figure 6: Samples of riverbeds generated with the algorithm.

ACKNOWLEDGEMENTS



The author is a scholar within Sub-measure 8.2.2 Regional Innovation Strategies, Measure 8.2 Transfer of knowledge, Priority VIII Regional human resources for the economy Human Capital Operational Programme co-financed by European Social Fund and state budget.

Adaptive Projection Displays: a low cost system for public interactivity

Jenna Dundas
Drexel University
3501 Market St.
Drexel STE 220
United States of America
(19104), Philadelphia,
Pennsylvania
jd574@drexel.edu

Michael Wagner
Drexel University
3501 Market St.
Drexel STE 220
United States of America
(19104), Philadelphia,
Pennsylvania
wagner@drexel.edu

ABSTRACT

Interactive digital public displays that track viewer's position are currently inaccessible to the average consumer. Many tracking systems available on the market are prohibitively expensive and are out of scope for small business owners to purchase. A system can be devised by researching and testing various consumer level tracking technologies in a low cost and accessible manner. Microsoft's Kinect in tandem with Unity3D offers a system that is straightforward to use and allows for ease of implementation. The resulting technique can be quickly carried out to create an interactive digital public display.

Keywords

Microsoft Kinect, projection, Unity3D, low cost, public display, interactivity, computer graphics

1 INTRODUCTION

The world is growing increasingly interactive, and displays that were once analog are becoming digital. Billboards are transitioning from paper to large-scale digital displays, which allow multiple advertisements on billboard. Within these new digital displays, the ability to switch and transform the visuals is as easy as a press of the button [Bor00b]. As these surfaces are transition from analog to digital the prospect for interactive public displays, become abundantly possible. These digital public displays have the capability to develop beyond just images and words to create a lasting impression on the consumer. This development of content on these public display systems is evolving, however, developers have yet to leverage today's technology [Alt01a]. Digital public displays can become more than just digital billboards on the side of the road. These displays can be signs in storefronts, advertisements in transportation centers, and museum displays.

Although public displays are making the transition from analog to digital displays, advertisers have not

moved beyond using these display areas as more than digital posters [Bor00b]. The full potential of these digital public displays is still an open area to explore. With the growing number of these displays developing within outdoor and indoor spaces, there is lies an ability to create new experiences, which change the framework of existing public displays [Ren00a]. While public displays can further explore interactivity within digital systems, there are existing ideas of interactive systems that the public know. For example after user notices that the public display is interactive, the display should react instantly to the user, to express usability [Mül00a].

With the increased number of digital public displays growing, low cost system can be created to track the viewer and create the illusion of depth and space. Interactive three-dimensional worlds can go unexplored because of the complexity of the communication that takes place [Hac00a]. Although Cursan was talking about the interaction between digital three-dimensional object and a user, this problem manifests itself within the public display system. Previously the ability to use three-dimensional cameras was only for companies that had a large budget. However with Microsoft's introduction of the Kinect, the ability for someone to implement three-dimensional cameras grew higher than ever before.

As a result of the availability of three-dimensional cameras a low cost system can be built, that allows small business and consumers to create interactive

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

three-dimensional public displays. The system is using a combination of off the shelf technology and low cost graphics engine to create an easily implemented interactive system. The system will allow producers to create displays that utilize three-dimensional graphics, light, and shadow to create the illusion of depth and space. Through this system, the three-dimensional camera will track the viewer's position and update the visuals depending on where the viewer is in relation to the display.

2 RECENT SOLUTION

Large-scale public displays have the freedom to transform the spaces around the world. As a result of the transition from analog displays to digital displays the ability to transform the content on the display allows for a number of possibilities [Bor00b]. The exploration of existing interactive displays and viewer tracking projects yields the possibilities of low-cost interactive system that explores three-dimensional graphics and light and shadow to achieve depth and dimension.

Interactive Large Scale Public Displays

Not all media facades are interactive some might transform a building, a space, or a wall in for the form of an advertisement [Fis00a]. Although these displays tend to not be interactive, researchers created a system, to explore large-scale media interaction on the facades of buildings. The project SMSlinsight transformed the building into a large interactive display, which received messages from the custom input device [Fis00a]. The display researchers built reacted to the input device rather than tracking the viewer's movement. However, they found that the users enjoyed having authority over the public display and being able to contribute to the visuals [Fis00a].

To explore the availability of low-cost technology, researchers focused on creating an interactive system that allowed many members to interact with the system. Through their exploration, they used three easily accessible pieces of technology. The researchers wanted the set up to be simple and not a complex gaming console, which resulted in their use of a camera, projector, and computer [Ozt00a]. With the technology, the researchers used multiple-human tracking algorithms to create an experience that predicts the future steps of various people, within the given camera area [Ozt00a]. Through their system, the ability to create a low-cost system for public interaction can be built. However, their system only details with two-dimensional images, mainly footsteps.

Conveying interactivity of the system can be one of the more troubled aspects of interactive public displays. Researchers found that an interactive system that through the use of mirror images of the participants

the recognized interactivity faster than a call to action did [Mül00a]. To create the shadow of the public viewer, researchers developed a system that used Microsoft's Kinect to track the viewer and create their shadow. Through this system, they found that an interactive system should convey interactivity and instantaneous communication, or the system will fail [Mül00a]. Public interactive systems have the potential to be engaging and fun. However, these systems need to take into account how they are calling the viewers to communicate with the system. As result researchers found that systems could be interactive as soon as someone comes into contact with the display, are the most effective and attention grabbing. The Focus was mainly on digital representations of the viewers, and not on creating three-dimensional worlds that are affected by the viewer and how the viewer might perceive this.

Digital Displays that Utilize Tracking Technology to Engage an Audience

Stookie is an interactive animated artwork that utilized three-dimensional projection mapping techniques [Jun00a]. Researchers on this project produced a head statue, which was built in a neutral facial position. This neutral position allowed them to then exploit three-dimensional projection mapping, to change the facial expressions and eye movement of the statues [Jun00a]. Through use of the Kinect, the researchers were able to create an interactive three-dimensional mapped experience for viewers. This system would update the eye position of the statue based on the movement of the public, and would call different animation states to convey emotions.

With consumer tracking technology, researchers created a system called FRAGWRAP. Through this system developed, researcher projected images on to a bubble that was infused with fragrance. As a result, they created the concept of scent mapping, which maps scent and visual information onto a physical object [Kyo00a]. Within this system, the researchers developed a method of tracking the position of the bubble and updating the projection mapped image to accompany the bubbles movement. By using the Kinect from Microsoft, they were able to use the RGB and depth camera to project images on the bubble and modify them in real-time [Kyo00a], and unity3D for the graphics. In addition to the Kinect system, they used Processing to implement form and location control [Kyo00a]. The ability to track objects and use projection mapping to display visual data is realized, through this project.

SnowGlobe is a spherical display that focused on creating an interaction with three-dimensional objects [Bol00a]. Through this display researchers successfully tracked the head of the viewer using a Vicon

motion capture device, this data was used to maintain motion parallax of the three-dimensional object in the display [Bol00a]. To track the head of the viewer researchers created custom glasses, which had three reflective infrared markers to track the location of the head [Bol00a]. Through this product, the researchers were able to exploit the tracked data and use it to create the parallax in the three-dimensional image. Additionally within this system the user had the ability to interact with the three-dimensional model that was displayed on the spherical surfaces. The researchers focused on creating a digital object that provided for one to completely walk around it, with the use three-dimensional projectors. The projection display was a spherical object that reflected a projected data off the hemispherical mirror, this allowed for a seamless curvilinear display surface [Bol00a]. This experiment in tracking with spherical display, show how tracking technology can be incorporated within in a public interactive system that explores light and shadow to create the illusion of depth and space.

3 PROPOSED SOLUTIONS

Currently, there is software and hardware that tracks people based on a number of different factors. Through the exploitation of existing software and hardware, the system can be devised that will update the projected artwork and create a visual parallax between the viewer and the projected display. A low cost responsive public display system can be created that will adjust depending on the viewer's perspective by utilizing off the shelf technology. The user's experience will change depending on the place the viewer, in relation to the display. The projected images will be adjusted depending on the viewer's position to achieve appropriate visual parallax for the viewer. The projected images will give the illusion that the digital imagery projected is behind the surface structure. Through the visuals depth and dimension will be exploited to fabricate a lasting impression on the viewer.

The proposed interactive system can be operational, utilizing four different components. Each of the four components of the system is easily accessible to obtain, and foster a low budget. The four components are tracking hardware, a graphics engine/scripts, computer, and a display unit. To built the responsive interactive system that reacts to the viewer's movement, one can utilize the combination of these four elements.

4 EXPERIMENTAL RESULTS

By researching into various tracking displays, and interactive public display system a select few products were tested further. These products were webcams, Microsoft's Kinect, and Vicon's Motion Capture cameras.

The three devices were evaluated on accessibility, volume of tracking area, and portability. Through the exploration of these products, the best product in the three categories was implemented into the end system.

Test of Off Shelf Technology

With web cameras, the tracking capability was remarkably good. However through the system of using FaceAPI and Unity the tracking area the viewer could be in was limited to very close to the camera area. The tracking area only included a less than a foot away from the camera and to the left and right. Additionally through this set up the system's graphics would become jumpy and jittery in comparison to other product's results. Although the tracking area was minimal in comparison to other products, the webcam is cheap and portable. The tracking area within the webcam was this technology's downfall, to be incorporated into the system.

Vicon's motion capture cameras are one of the best three-dimensional cameras currently on the market today. However, this product is expensive and not viable available product, to the average consumer or small business. As a result of the price and accessibility the Vicon's camera, their product was not an option for a low-cost interactive system.

Within the last few year's one consumer product has revolutionized gaming, interactivity, and interfaces around the world. Microsoft's Kinect is one of the first consumer level three-dimensional cameras that use infrared light and sensors to gain the location of the viewer and several other positional data from them. This camera is not only affordable to consumers and small businesses, it is easily available and reasonable in price, if someone would like to buy one. Microsoft's Kinect can hang, place, or move easily in comparison to other three-dimensional cameras. They are lightweight, small, and compact. As a result of this, they can easily be packed and transported to various locations. Additionally this camera gives a larger tracking volume than the area a webcam could achieve. As a result of these factors, the Microsoft's Kinect was the choice product to use within this system.

Prototype System

The system contains four elements a tracking device, a computer, a graphics engine, and a display unit. To build a system with a low budget, the rough system respectively used Microsoft's Kinect, a free version of Unity3D, Microsoft's SDK, a free script package called Kinect with MS-SDK, a few custom script, a laptop, and a projector. The final cost to gather these factors vary depending on the type and quality projection display system is employed and the availability of a laptop

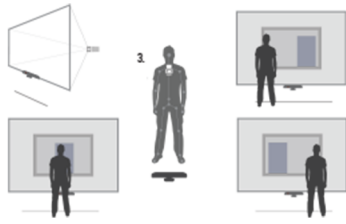


Figure 1: Prototype display process

computer. However, the cost can slide between 400 dollars to 1000 dollars.

The rough system tracks a user using Microsoft's Kinect infrared emitter, and sensor. Through this infrared light bounces off the viewer to locate the position of their body to the Kinect. Once the position of the viewer is found through the Kinect. The skeleton data is brought into Unity3D using Microsoft's software development kit and a script package that bridges the software development kit and Unity3D. Once this data is brought into the game engine, a custom script is applied to the camera. This script takes the place of the viewer's center chest, and updates whenever the viewer moves in relationship to the visuals. The viewer can move forward, backward, right, and left and the visuals will update with appropriate parallax depending on their position. Additionally another addition is the resetting of the scene once the viewer has left the volume of the Kinect tracking capabilities. Through the use of unity3D one can use three-dimensional graphics, light, and shadow to fake the illusion of depth and space in a location that does not have any.

Graphics Engine

Through this interactive art piece, the goal was to create a design that could be easily replicated for anyone to incorporate within his or her own particular projects. As a result of this the project focused on finding an existing method for graphics, rather than trying to write a custom graphics engine. With these features in mind, one option came to mind. This graphics engine is called Unity3D, which is commonly used as a gaming engine. However, the program is robust and allows for the convergence of different projects and ideas that go beyond a gaming environment. Through this product, the ability to create real-time graphics became straightforward, and simple to be incorporated into an interactive art piece.

The exploitation of this existing product allows for a designer to utilize existing features of the program. For example, the ability to light the scene and use several different rendering techniques have been created already, and can be easily incorporated to create a higher quality image than one might be able to do when working on their own person graphics engine product. One

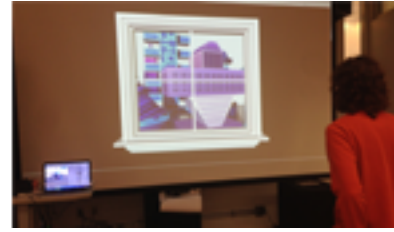


Figure 2: Prototype System

of the downfalls of this graphics engine is that some results are only available within the pro version of the product. However, the completely free version, that anyone can download and use, is one of the best things about this product. Essentially the free version of the engine is the same as the professional version, just with a handful of features not included. Features that are included in the professional copy and not within the free copy are fog, rendering effects, different types of shadows, and image effects. However within the free version the designer can only use hard shadows within a directional light, which gives the designer some freedom. However, this freedom does not compare to the flexibility to create effects within the professional version.

Scripts

To work with the Kinect within unity3D, there is a small set up that one needs to do to start working with the product. Within this interactive art piece, the project relied on a few different components. For example the Microsoft software development kit (SDK), the script that bridges communication between the SDK and Unity3D, and a custom script.

Microsoft's Kinect Software Development Kit

Through Microsoft's website, they offer a software development kit also known as an SDK for free download on their website. This SDK is only for windows based machines, and uses C++, C#, or visual basic to create Kinect based applications. Through this SDK, the data from the Kinect is brought to the computer. Additionally through this SDK the drivers for the device are installed, this way the Kinect can communicate to the computer.

Kinect with MS-SDK

Through the Unity asset store, there is a free package that bridges the Microsoft SDK to Unity3D. This way someone can use the skeletal data, depth data, gesture, and other information the Kinect offers to explore. This method was the best way to bring this data into Unity in comparison to other methods, and other various scripts available on the Internet. Through this package, there are a number of scenes that explore different aspects of data found from the Kinect. Through this interactive

art piece the use of the pointman script was the main source of information for the position of the viewer.

Custom Script

Although Microsoft's SDK and the Unity3D extension allow for skeletal data to be brought into the graphics engine, there is still one last step. This last step is a small script that is applied to the camera within a given Unity3D scene, that updates the camera site depending on the viewer's position. What this script does is take the position of the of the placed pointman center shoulder joint and translates this position to the camera position. Through this script the camera gets x, y, and z values for its position within a given scene. By giving the camera x, y, and z positions this allows the viewer to move left, right, forward, backward, up, and down with the visuals updating accordingly.

Additionally to the giving the camera movement through the body position of the viewer the script can increase or decrease the amount of movement the camera does is response to the viewers movement. This allows for a designer to decide how much or how little their designs will react to the viewers of their final designs.

This script also gives the designer the option to put the distance the Kinect is in front or behind the display screen and have there visuals be pushed back in space depending on the inches the give. To accomplish this, the designer must create their visual assets and then combine them into a unity prefab, and place it within their resources folder. Once the prefab is created a designer can drag the prefab into the given place within the unity script. The script takes the given inches the screen is and converts it to meters that are the starting unit of the graphics engine. After it converts the inches to meters it places the artwork at that given spot in the three-dimensional space.

Environmental Issues

Microsoft's Kinect the display system can be easily implemented as a low cost interactive public display system. However, within the confines of using a Kinect and a projection based display system the best environment for the system is within indoor or low light environments. While using projectors allows the producer to create a large display image cheaper than buying a large LCD monitor, the visual quality of the image diminishes when light floods the projection area.

Applications

With digital displays increasing with numbers throughout everyday life, the ability to create interactive displays is growing with the amount of displays transitioning from analog to digital. The process created for this

interactive art piece can express how digital public displays can utilize existing technology to create the illusion of depth and space. The set-up of this art piece explores the use of light, motion, and space to create the illusion of depth where there is none in the actual display. This effect is created by tracking the viewer's location and updating visuals depending on their place within the space in relation to the display system. Three areas retail, advertisements, and historical reconstructions can explore the use of this system further.

Retail

The resulting idea can be applied to retail conditions in two different manners. The first manner ins which this process of creating and interactive display can be used to create a look and feel for the store logo, within their storefront. The interactive display can take the form of a logo in front of visuals that match the store's design image. Through this the logo can be in front of the graphics and appear as though it is jumping off the front of the digital display setting.

Secondly retail can utilize this process by allowing people to visualize the product within various different conditions. For example, a consumer can have a picture of their living room and place a particular product they are thinking of purchasing within this three-dimensional world. Unlike other retail options, the consumer and experience it on a visual and spatial level. Through this system the graphics can promote a product, an image, or a feeling that the designer is trying to convey to the users.

Advertisements

One encounters a number of different displays, which range from digital ads on computers to ads on the street or transportation system. Through utilizing this tracking framework for ads information can pop out to the consumer that the marketer would like to emphasise, rather than relying solely on visual structure of the graphic design to push their marketing message to the consumer. Through the use of the three-dimensional effect brand names and core information can be pushed forward in the space emphasis importance to the viewer, where the less important information can be pushed visually back in space.

Digital Reconstructions

Through the formation of this interactive art piece, the use of the process can be applied to historical digital reconstructions. This digital diorama can be placed within museums to allow visitors to experience worlds, which they would not able to experience without the digital reconstruction. The advantages to the digital version of this versus the recreation of the situation in real life is the ability to the museum to make updates

and changes to situations that might change depending on new information arises about the history of the situation or artifact.

5 CONCLUSION

Through the formation of this interactive public display system, the design will be actualized to update three-dimensional images based on viewer's location. The interactive system will focus on fostering a low-budget, in hopes of allowing anyone to create a three-dimensional interactive public display system. Through using products like Microsoft's Kinect and Unity3D, this method is accessible to the general public. The interactive system has applications in advertisements for small business, art exhibitors, and historical reproductions of environment. Through this interactive system, the visuals will be more interactive and engaging than a poster with visuals. The interactive system will actively mediate the viewer's world, and allow for digital displays to become three-dimensional rather than digital poster displays. The system will utilize three-dimensional graphics, light, and shadow to create the illusion of depth and dimension on a flat surface. Through this system, an immediate reaction to the viewer will allow the display to communicate its interactivity to the public. As a result, this will drive attention to the display and potentially to the store or exhibit.

6 REFERENCES

- [Alt01a] Alt, F., Schmidt, A., and Müller, J. Advertising on public display networks. *Computer*, 45(5):50-56, 2012.
- [Bol00a] Bolton, J., Kim, K., and Vertegaal, R. Snowglobe: a spherical fish-tank vr display. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 1159-1164. ACM, 2011.
- [Bor00b] Boring, S., and Baur, D. Making public displays interactive everywhere. *Computer Graphics and Applications*, IEEE, 33(2):28-36, 2013.
- [Fis00a] Fischer, P.T., Zollner, C., Hoffmann, T., Piatza, S., and Hornecker, E. Beyond information and utility: Transforming public spaces with media facades. *Computer Graphics and Applications*, IEEE, 33(2):38-46, 2013.
- [Hac00a] Hachet, M., De La Rivière, J.B., Laviole, J., Cohé, A., and Cursan, S. Touch-based interfaces for interacting with 3d content in public exhibitions. *Computer Graphics and Applications*, IEEE, 33(2):80-85, 2013.
- [Jun00a] Jung-ah, I., Semi, K., Jaeyoung, K., Bonhwa, C., Hwanik, J., Junghwan, S. Stookie. In *SIG-GRAPH Asia 2012 Posters*, page 38. ACM, 2012.
- [Kyo00a] Kyono, Y., Yonezawa, T., Nozaki, H., Nakasawa, J., and Tokuda, H. Fragwrap: fragrance-encapsulated and projected soap bubble for scentmapping. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing* adjunct publication, pages 311-314. ACM, 2013.
- [Mül00a] Müller, J., Walter, R., Bailly, G., Nischt, M., and Alt, F. Looking glass: a field study on noticing interactivity of a shop window. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 297-306. ACM, 2012.
- [Ozt00a] Ozturk, O., Matsunami, T., Suzuki, Y., Yamasaki, T., and Aizawa, K. Real-time tracking of humans and visualization of their future footsteps in public indoor environments. *Multimedia Tools and Applications*, 59(1):65-88, 2012.
- [Ren00a] Ren, G., Li, C., O'Neill, E., and Willis, P. d free-hand gestural navigation for interactive public displays. *IEEE Computer Graphics and Applications*, 33(2):47-55, 2013.

Connect-S: A Physical Visualization Through Tangible Interaction

Ken Giang

Department of Industrial Design
Eindhoven University of Technology
Eindhoven, The Netherlands
k.giang@student.tue.nl

Mathias Funk

Department of Industrial Design
Eindhoven University of Technology
Eindhoven, The Netherlands
m.funk@tue.nl

ABSTRACT

In our current society, open data streams are more and more available through the Internet. This data can have an increasing impact on everyday life. Its full potential can, however, only be reached through better integration and new interfaces. The goal of this project is to explore the possibilities of repurposing public information in a developing area of a large city in the Netherlands. Can we create a tangible interaction with use of physical visualization of these data streams? A series of prototypes have been made to develop a physical visualization through the method of research through design. Users were involved in expert panels and interviews to fine-tune and create a final prototype, Connect-S. The concept shows the opportunities of using physical visualization in connection with physical interaction for browsing and navigation.

Keywords

Physical visualization, tangible interaction, interactive systems, gesture control

1. INTRODUCTION

Every day of our lives, data is gathered about our activities and interactions with people, objects and systems. Technology is advancing at a rate where user generated data can be implemented in systems to make life more efficient and more connected. A lot of data sets are now publicly available through systems and services such as Xively (formerly known as Cosm)[Xiv13], but also through more mediated interfaces embedded in social networks and mobile apps (e.g., location-based weather forecast, media recommendations etc.).

Can sources of raw data be turned into meaningful information, which could fit our daily life better? [LW01] This research project focuses on a future scenario for a former industrial, but now developing area of the municipality of a large city in the Netherlands. Within this area, different types of sensors will be placed and open data streams are available in the near future. In addition, other external data sources will be linked. These open data

streams can be used to improve information distributed about the area to inhabitants, visitors and professionals in the area.

Adopting the methodology of research through design [ZFE07], different concepts have been developed to create both a physical visualization of open data streams as well as means to interact with the data.

The purpose of this paper is to address the potential of open data streams and the application in everyday life through accessible and easy-to-use tangible interfaces. In the following, after outlining related work, a series of prototypes leading to a final concept are presented, explored and evaluated with involvement of stakeholders, experts and potential users. The paper concludes with a summary and an outlook on future opportunities.

2. RELATED WORK

Inspired by the earlier drawings and concept sketches of the architects envisioning the city area in 2020 [FPW09], a benchmark performed on existing work related to either public installations [BD08, RKM10] or systems connected to the web [BT10, KBM02].

While these projects were mainly public installations or visions of future city areas [SKP11], installations or visualizations of open data streams in the context of the Internet of Things [FPC09] at the time did not implement system elements in them with the

□ Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

intended interactivity and appropriate use of information.



Fig 1. Existing project involving projection mapping on a scale model as visualization [BS11].

Aside from the concept of using data visualization in public, many projects embed techniques such as projection mapping, augmented reality and the use of microcontrollers and sensors to enhance the experience [BKP01].

3. Connect-S

The system that is developed to address the challenge of fitting data to everyday use is called Connect-S. It is a publicly accessible visualization in form of table with a scale model of the city area, showing open data streams and public information through layered animations projected from the bottom of the table onto the scale model. Passersby could interact with this data related to the surrounding by gesture recognition, and based on the distance between the hand and the table, navigation through layers of information was made possible. These layers of information were categorized into social media feeds, traffic and public transportation, event and planned activities in the area, and local activity of citizens.

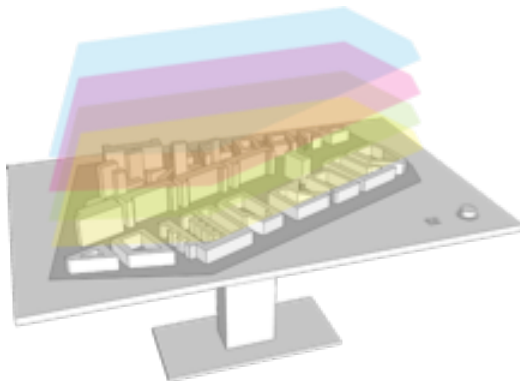


Fig 2. Sketch of the Connect-S table, including visual representations of categorized data layers.

Research through design

Based on the research through design approach, an iterative process was adopted to create this physical visualization. Starting with a scale model of the city area and the use of projection mapping as a method to animate and visualize the data projected directly

onto the scale model, several explorations were made and discussed with focus groups, experts and inhabitants of the area in order to fine-tune the animation and the interaction with the proposed concept.

The method allowed for a process of refining the interaction and the implementation of technology throughout several phases, and through a set of prototypes, several possibilities of information representation and visualization were built and tested. Another prototype focused on the user interaction with the table and explored options to navigate and interact with different, overlapping layers of information by using a combination of ultrasonic sensors and direct controls such as rotary controllers and buttons.

These explorations resulted in the final prototype, which was optimized by user involvement and designed to show the possibility of a public interactive system based on physical visualization of data, as shown in the following section.

Experience Prototype

The final prototype is a moderately scaled down version of the drawing shown in Fig. 1. This high fidelity prototype serves as an experience prototype to show the potential of combining physical visualization and tangible interaction.

The prototype consists of a laser-cut wooden scale model of the city area map (including buildings and roads, cf. Fig. 5) with an integrated LCD display underneath the surface to visualize different information layers.

Users could navigate through the connected data by either changing the layers of information through hovering their hand on different heights above the scale model (cf. Fig. 3).



Fig 3. User testing the experience prototype

Another interaction possibility was the use of a dedicated rotary controller to skim through time. This implementation was needed to show for example planned activities in the area over a period of time, but also to look at a range of public transportation options.

The layers of information consisted of the following categories:

- *Social Media layer*, consisting of messages, photographs and videos taken in the area and displayed on the corresponding location. Information that is made public on the web is repurposed as informative data to visitors and inhabitants of the area.
- *Traffic Information layer*, sets of information related public transportation and related information to traffic such as the available parking places and traffic jams. The public transportation would be displayed as a dynamic logo to inform users of the transportation schedule.
- *Events layer*, a more scripted layer of information. Organizations and companies could use this layer to promote and announce activities and events. By introducing a separate layer for summarizing events and activities users can easily foresee (future) events. This information would also be mapped out on the scale model.

- *Intensity layer*, information created by users and showing hotspots of areas based on sensors and information that is made public over the Internet. The use of these data streams and information the concept can be considered as a physical visualization of a 5-dimensional model of values: $(x, y, z, time, intensity)$

With use of these five dimensions, the concept would be able to use various in- and outputs to function. The X and Y dimensions are the dimensions that are used for to display location; the values are translated to a pixel grid on a LCD display. The Z dimension is controllable and mapped to the ultrasonic sensor, measuring the distance of the hand and showing one of the layered animations. The Time dimension would be mapped to a different controller; a rotary dial fulfilled the purpose of scanning in the past, present and future by browsing through different animations per layer. Intensity is only shown on the display itself. By creating points and a surface surrounding these points on the scale model, specific locations and sensors could be visualized representing dynamic content.

Based on the 5-dimensional model and the use of these four layers of information the interaction was fine-tuned appropriately with through the use of an LCD display, ultrasonic sensor and a rotary dial (cf. Fig. 4). The action possibilities of the experience prototype were influenced considerably by the practicalities of a public system and efficiency in mind.



Fig 4. Experience prototype with scale model and integrated LCD display, controllers (ultrasonic sensor and rotary dial) on the left.

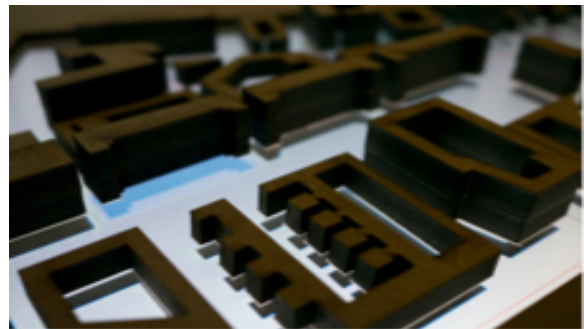


Fig 5. Close up view of the scale model and animation on the integrated LCD display

4. USER EXPLORATION

Within the design process of the Connect-S table, several methods were applied to validate the final concept and to receive feedback from users:

Expert panel

An expert panel was used to evaluate the current concepts and at the same to map out the design implications for the context. Through a series of informal interviews with experts in the field of architecture, organizations in the area and inhabitants, information was gathered about their vision and opinion of this plan for the future.

The findings from these interviews were used to generate concepts in the explorative phase of this research. Furthermore, requirements and practicalities for public installations and useful systems could be mapped based on this expertise and user involvement.

User confrontation

Different users and experts were asked for an informal feedback session about the final Connect-S prototype, in which they could experience local data in a spatial visualization by interacting with the table.

Findings from these sessions proved to be useful to confirm and fine-tune the general impression of the concept. While the prototype was tested using a few scenarios of using public data, the involvement of users proved itself to be useful even on this scale of testing the concept with an experience prototype.

5. DISCUSSION

Connect-S showed that the use of (moderate levels of) tangible interaction in a physicalized visualization can be useful as a medium to display and interact with intangible data that is collected and publicly available via the Internet, but not accessible for everyday use by inexperienced stakeholders. Through the design process and the inclusion of users during development, a physical visualization based on the information about the city area was developed, which would enable users to use information more actively and efficient in their everyday life similar to other physical visualizations [JDF13].

In comparison with earlier studies and projects involving public installations and implementation of technology in city areas [Sta11], this project did not only address the aesthetics and implementation of a public system, but also emphasizes opportunities of using tangible interaction as a means to include open data in a meaningful way.

One of the main reasons for choosing deliberately for a solution, which makes the interaction tangible, was due to the target user group. The Connect-S concept is focused on different users such as inhabitants of the area, visitors and working professionals. By placing multiple tables in the most populated and frequently visited areas users can interact and use this public installation both as a visualization of the area as well as a functional system strengthening the unique identity of the area with modern technology.

This project was developed and evaluated with the involvement of an expert panel and informal interviews. The concept was received positively by all parties, but needs to be tested on a larger scale in the future to make stronger conclusions to strengthen the use of open data for tangible interfaces. Also, given the quite generic implementation and mapping of data to visualization and interaction, the concept could be easily scaled up to different parts of the same or different cities. As a starting point, however, designing for a specific context, such as the future city area proved to give the research a more practical point of view, thus affecting the outcomes such as the experience prototype that users and stakeholders could directly and actively test and give feedback on. A strong and realistic context certainly accelerates the process of designing and developing physical visualizations with animated data layers and tangible interaction. Through this approach, the table goes beyond the boundaries of a (physical) data visualization that only shows the aesthetics of data, but highly constrains the amount of information presented.

6. FUTURE WORK

Some pointers for the future have been mentioned in the previous sections: The Connect-S concept

inspires different projects related to tangible interaction through the use of visualization of data.

Developing a prototype with layered animations and gesture control is one of the possibilities to create a tangible interaction. While this seemed the appropriate mapping for the context in this paper, other methods of mapping would be possible depending on the design requirements. This new challenge of system development arises with the freedom of mapping the data to an appropriate visualization and interaction without the boundaries of the physical form. An example is AMP (cf. Fig 6), a lamp that uses an original method of projection to display the activity around the city by means of social media streams. The concept shows different interaction possibilities in both the use of visualizations and the method of interaction by its user in comparison with this research project [ET13].



Fig 6. AMP concept projecting a city map and displaying current activity based on data sourced from social media.

This paper introduces an approach for developing concepts concerning visualization of data and the tangible interaction for a specific context and user group through the method of research through design [ZFE07]. For future work, one can consider developing an application of data visualization using the methodology to create real life applications of this field of research involving stakeholders and future users.

7. CONCLUSION

In this paper Connect-S is introduced, an interactive scale model of a city area with animated data layers.

A tangible interaction was presented as a means to give purpose to the open data streams and the physical visualization of information related to the city area. While the interaction with Connect-S has been not explored extensively over a period of time, this ongoing research shows potential of integrating more systems in daily life that inform and give back user generated data.

The visualization of data can be used as a medium to communicate information from open data streams. Embedding this information into systems in daily life can enrich the experience and the use of data on

manners that have not been explored in the past. Public systems as in the case of Connect-S show that technology can be integrated on a new level of interaction, a level that invites users to interact and inform on a daily basis and can grow over time depending on the use of the system.

8. REFERENCES

- [BD08] Boi, Clara; Diaz, Diego, The Hybrid City: Augmented Reality for Interactive Artworks in the Public Space. In *The Art & Sci. of Interface & Inter. Des, SCI 141*, pp. 141–161, 2008.
- [BKP01] Billinghamurst, Mark; Kato, Hirokazu; Poupyrev, Ivan, Collaboration with Tangible Augmented Reality Interfaces. In *HCI International 2001*, New Orleans, LA, 2001.
- [BS11] Beranek, Roman; Staub, Jonas, Showroom Richti Areal, 11-April-2011. [Online] Available: <http://www.projektil.ch/portfolio/showroom-richti-areal/>. [Accessed: 31-Mar-2014]
- [BT10] Boussard, Mathieu; Thebault, Pierrick: Navigating the Web of Things: Visualizing and Interacting with Web Enabled Objects. In *ISoLA 2010, Part I, LNCS 6415*, pp. 390–398, 2010.
- [ET13] Eggen, Berry; Terken, Jacques, Eindhoven designs: Volume seven, ISBN: 978-90-386-3489-0, pp. 6-11, 2013.
- [FPC09] Funk, Mathias; Putten, Piet van der; Corporaal, Henk, Analytics for the Internet of Things, in *Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 4195–4200, 2009.
- [FPW09] FPW Eindhoven, The Creative City: STRP-S [Online], Available: http://www.strijp-s.nl/App/Resources/media/documenten/Strijp-S_brochure_UK.pdf [Accessed: 31-Mar-2014]
- [JDF13] Jansen, Yvonne; Dragicevic, Pierre; Fekete, Jean-Daniel, Evaluating the efficiency of physical visualizations, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2593–2602, 2013.
- [KBM02] Kindberg, Tim; Barton, John; Morgan, Jeff; Becker, Gene, et al, People, Places, Things: Web Presence for the Real World. In *Mobile Networks and Applications 7*, pp. 365–376, 2002.
- [LW01] Lange, Michiel de; Waal, Martijn de, VIRTUEEL PLATFORM RESEARCH: Ownership in the hybrid city. [Online] Available: <http://virtueelplatform.nl/english/publications> [Accessed: 31-Mar-2014].
- [RKM10] Roalter, Luis; Kranz, Matthias; Moller, Andres, A Middleware for Intelligent Environments and the Internet of Things. In *UIC 2010, LNCS 6406*, pp. 267–281, 2010.
- [Sta11] Stalder, Ursula, Digital Out-of-Home Media: Means and Effects of Digital Media in Public Space. In *Pervasive Advertising, Human-Computer Interaction Series*, pp 31-54, 2011.
- [SKP11] Schaffers, Hans; Komninos, Nicos; Pallot, Marc; Trousse, Brigitte; Nilsson, Michael; Oliveira, Alvaro, Smart Cities and the Future Internet: Towards Cooperation Frameworks for Open Innovation. In *Future Internet Assembly, LNCS 6656*, pp. 431–446, 2011.
- [Xiv13] Xively.com, *Product Page*, 14-May-2013. [Online]. Available: <http://www.xively.com/>. [Accessed: 31-Mar-2014].
- [ZFE07] Zimmerman, John; Forlizzi, Jodi; Evenson, Shelley, Research Through Design as a Method for Interaction Design Research in HCI *Human-Computer Interaction Institute*. Paper 41, 2007.

Real time vehicle detection and tracking on multiple lanes

Kristian Kovačić, Edouard Ivanjko, Hrvoje Gold
 Department of Intelligent Transportation Systems
 Faculty of Transport and Traffic Sciences, University of Zagreb
 Vukelićeva 4, HR-10000 Zagreb, Croatia
 kkovacic@fpz.hr, edouard.ivanjko@fpz.hr, hrvoje.gold@fpz.hr

ABSTRACT

Development of computing power and cheap video cameras enabled today's traffic management systems to include more cameras and computer vision applications for transportation system monitoring and control. Combined with image processing algorithms cameras are used as sensors to measure road traffic parameters like flow, origin-destination matrices, classify vehicles, etc. In this paper development of a system capable to measure traffic flow and estimate vehicle trajectories on multiple lanes using only one static camera is described. Vehicles are detected as moving objects using foreground and background image segmentation. Adjacent pixels in the moving objects image are grouped together and a weight factor based on cluster area, cluster overlapping area and distance between multiple clusters is computed to enable multiple moving object tracking. To ensure real time capabilities, image processing algorithm computation distribution between CPU and GPU is applied. Described system is tested using real traffic video footage obtained from Croatian highways.

Keywords

Multiple object detection, intelligent transportation system (ITS), vehicle detection, vehicle tracking, algorithm parallelization, trajectory estimation

1 INTRODUCTION

Video sensors or cameras combined with image processing algorithms are more and more becoming the approach to today's road traffic monitoring and control. They have become robust enough for continuous measurement of road traffic parameters [Con14]. From the obtained video footage high level traffic information can be extracted, i.e. incident detection, vehicle classification, origin-destination (OD) matrix estimation, etc. This information is crucial in advanced traffic management systems from the domain of intelligent transportation systems (ITS).

In order to provide high level traffic information using a computer vision system, vehicle detection has to be implemented first. Most often used current approaches are based on: (i) foreground / background (Fg/Bg) image segmentation methods where moving (foreground) objects are separated from static (background) objects as described in [Con12a] and [Con07]; (ii) optical flow computation of specific segments (moving clusters) in an image [Con00]; and (iii) vehicle detection

algorithms based on the Hough method [Con12b] or on Haar-like features [Con06]. For real time systems most suitable approach is the Fg/Bg image segmentation method because of its low computational demands. Low computational demand is important in creating real time systems especially for multiple object detection and tracking systems.

After a vehicle in an image has been detected, its movement is tracked by storing its pose (position and orientation) and its pose change for each time segment. Using the saved vehicle poses and pose changes, its trajectory can be estimated. Trajectory estimation can be performed using various approaches (mostly a prediction/correction framework with odometry as the motion model) and its basic task is to determine a mathematical function which will best describe given set of points (vehicle poses) in 2D space. Vehicle trajectory estimation can be separated into following three parts: (i) finding a function f which will best describe given set of points in 2D space; (ii) transformation of parameters from 2D to 3D space model; and (iii) computing movement parameters such as vehicle direction, velocity and acceleration/deacceleration in 3D space model [Pon07].

Typical commercial computer vision based traffic monitoring systems use one camera per lane to ensure accurate and robust traffic parameters measurement. This presents a drawback since many cameras are needed for roads with multiple lanes which makes such systems expensive. This paper tackles the mentioned problem

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

by modifying the image processing part to enable vehicle detection and tracking on multiple lanes in real time. Image processing is parallelized and its execution is distributed between the CPU and GPU. So, only one camera per road is needed making such systems simpler and cheaper.

This paper is organized as follows. Second section describes the approach used in this paper. Third section describes the vehicle tracking algorithm. Following fourth section presents obtained results. Paper ends with conclusion and future work description.

2 VEHICLE DETECTION

Basic work flow of the proposed system consists of four main parts: (i) image preprocessing; (ii) Fg/Bg image segmentation; (iii) pixel clusterization; and (iv) multiple object tracking. The task of the image preprocessing part is to enhance the image imported from a video stream using a blur filter. After preprocessing, the image is passed through Fg/Bg image segmentation to separate foreground (moving) from background (static) segments in the image. This method is based on creating a background model from a large number of preprocessed images and comparing it with the new preprocessed image. The Fg/Bg segmentation result is then passed through pixel clusterization which computes location of each object (vehicle) in a scene and tracks its trajectory through consecutive images (frames). Final part for multiple object tracking of the proposed system performs also vehicle counting using markers defined in the scene.

In order to accurately detect, track and count vehicles obtained traffic video needs to satisfy following requirements: (i) camera perspective in the video footage must be constant over time (fixed mounted camera); (ii) all moving objects in the scene are vehicles (system does not perform classification between detected moving objects); and (iii) traffic video must not be previously preprocessed with image enhancing algorithms (for example auto-contrast function which can degrade the system accuracy).

Algorithms for image downsampling, image preprocessing and Fg/Bg segmentation are executed entirely on GPU. Today's GPUs enable high parallelization support for mentioned algorithms and can reduce execution time for basic image processing operations. Part of the system related to image preprocessing is performed with a single render call. In Fg/Bg segmentation, creation of background model is performed with 8 render

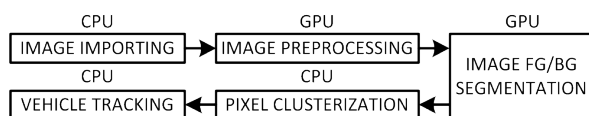


Figure 1: CPU/GPU computation distribution.

calls to process one frame. Moving object detection (comparison of background model and current image) is performed with one single render call. Algorithms for pixel clusterization and vehicle tracking are not suitable to run on GPU because of their structure and therefore are run entirely on CPU. Pixel clusterization and vehicle tracking algorithms are made of many dynamic nestings and IF clauses, and too complex to run on GPU in parallel. The mentioned image processing workflow and distribution of computations between the CPU and GPU is given in Fig. 1.

2.1 Image Preprocessing

Every image imported from the road traffic video footage contains a certain percentage of noise. Noise complicates the vehicle detection process and significantly reduces the accuracy of the described system so it needs to be minimized. For noise reduction, blur filters are commonly used. They reduce the number of details in the image including noise. In the proposed system a 4×4 matrix Gaussian blur filter is used for noise reduction. Work flow of image preprocessing is given in Fig. 2.

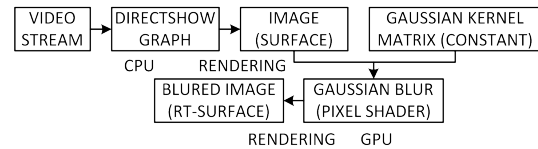


Figure 2: Image preprocessing work flow.

2.2 Foreground / Background Image Segmentation

After the imported image has been successfully preprocessed, Fg/Bg image segmentation is performed as shown in Fig. 3. As mentioned, this process consists of creating a background model of the scene and comparing computed background model with the latest image imported from the video [Con13]. The background model is obtained using the following equation:

$$BG_t = BG_{t-1} + \left[\frac{\sum_{i=1}^n \text{sign}(I_i - BG_{t-1})}{n} \right], \quad (1)$$

where BG_t represents value of a specific pixel in the background model for current frame, BG_{t-1} is value of a specific pixel in the background model for the previous frame, I_i is a value of the certain pixel in i^{th} image, and n is the number of stored images.

By comparing mentioned pixels in imported images, every pixel in the currently processed image can be classified. If difference between current image pixel

value and background model pixel value is larger than specified threshold constant, pixel is classified as a part of a foreground object. Otherwise it is considered as a part of the background model. Result of preprocessing and Fg/Bg image segmentation is given in Fig. 4.

2.3 Pixel Clusterization

After each pixel in the image is classified as part of a foreground object or as a segment of the background model, pixel clusterization needs to be performed. Used approach is based on marking all adjacent pixel that have the same pixel value as a part of a specific cluster. Afterward, all pixels within the same cluster are counted and their minimum and maximum values of x and y coordinates are found. With mentioned information clusters can be represented as rectangles. Rectangle center is used as the cluster center.

In the proposed system, pixel clusterization is performed only on foreground pixels. Additionally, all clusters that do not contain enough pixels related to them are discarded and excluded from further processing.

3 VEHICLE TRACKING

The clustering part returns a large number of clusters i.e. objects or possible vehicles. To sort out objects

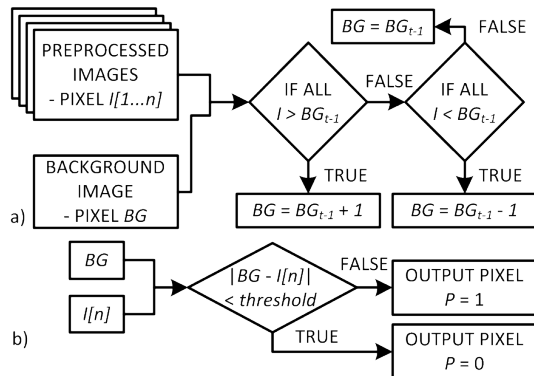


Figure 3: Fg/Bg image segmentation work flow: a) background model creation, and b) background model and current image comparison.

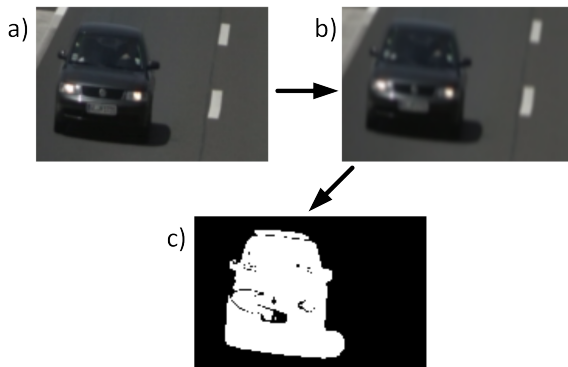


Figure 4: Original image (a) passed through preprocessing algorithm (b) and Fg/Bg segmentation (c).

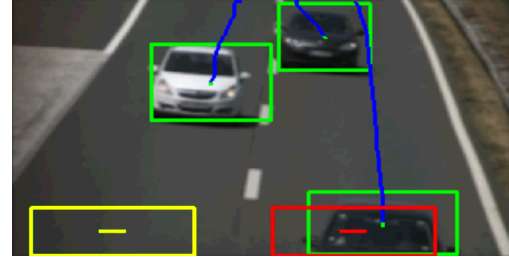


Figure 5: Vehicle tracking and counting on two lanes.

that are not vehicles, filtering is applied. In the proposed system, spatio-temporal tracking of objects in a scene is used for filtering. Every currently tracked object in the scene is compared with each cluster detected in the current image. Cluster that does not match with any of the previously detected objects is set as a new object. Cluster matching is performed by searching for the largest weight factor related to the cluster and specific object. Cluster will be assigned to the object with highest weight factor. Appropriate weight factor w is computed using the following equations:

$$w_{dist} = 1 - \frac{d - d_{min}}{d_{max} - d_{min}}, \quad (2)$$

$$w_{area} = 1 - \frac{a - a_{min}}{a_{max} - a_{min}}, \quad (3)$$

$$w_{cover} = \frac{a_{is}}{\max(a_{obj}, a_{cl})}, \quad (4)$$

$$w = \frac{w_{dist} + w_{area} + w_{cover}}{3}, \quad (5)$$

where d is distance between location of the specific cluster and estimated object location, d_{min} and d_{max} are minimum and maximum distance between all clusters and processed object, a is difference between the cluster area (size) and estimated object area, a_{min} and a_{max} are minimum and maximum difference between all clusters area and estimated object area respectively, a_{is} is intersection area between cluster and object, a_{obj} is area of the object, and a_{cl} is the processed cluster area.

To compute the distance between location of the specific cluster and estimated object location their geometric centers are used. Cluster and object area are computed as their surrounding bounding box area.

4 EXPERIMENTAL RESULTS

The proposed system has been tested using real world traffic footage captured on a highway with two lanes near the city of Zagreb in Croatia. Camera was mounted above the highway and passing vehicles were recorded using a top view camera perspective as given in Fig. 5. Duration of the test video was 10 [min]. Obtained original video resolution is 1920×1080 pixels (RGB).

Approach		Vehicle count		
		Total	Lane	
			Left	Right
Overlap check	Hits	126	65	61
	FP / FN	0/6	0/5	0/1
	Accuracy	95.6%	92.9%	98.4%
Trajectory check	Hits	129	68	61
	FP / FN	1/4	0/3	1/1
	Accuracy	96.2%	95.8%	96.8%
True vehicle count		132	70	62

Table 1: Counting results of the proposed system.

For experimental results, two approaches for vehicle counting were tested. Both are based on markers (virtual vehicle detectors). Markers are placed in bottom part of the scene on each lane as shown in Fig. 5 with yellow and red rectangles. Yellow color denotes an inactive marker and red color an activated marker. Edges of markers are perpendicular to the image x and y axis. When a vehicle passes through marker and a hit is detected, counter for that marker is incremented. First approach checks if an object is passing through marker with its trajectory and second approach performs check if an intersection between marker and object exists. Both approaches discard all objects whose trajectory direction is outside of a specific interval. In performed test, all moving objects need to have their direction between $90 - 270$ [°] in order not to be discarded. Objects also need to be in the scene for more than 30 frames. Value of the threshold constant used in Fg/Bg segmentation method is 10 and number of consecutive images used when creating background model (n) is 105. Blue lines in Fig. 5 represent computed vehicle trajectory. Experimental results are given in Tab. 1. FP represents false positive and FN represents false negative hits. True vehicle count is acquired by manually counting all passed vehicles.

In Fig. 6, execution time is given for various resolutions tested on Windows 7 (64bit) computer with

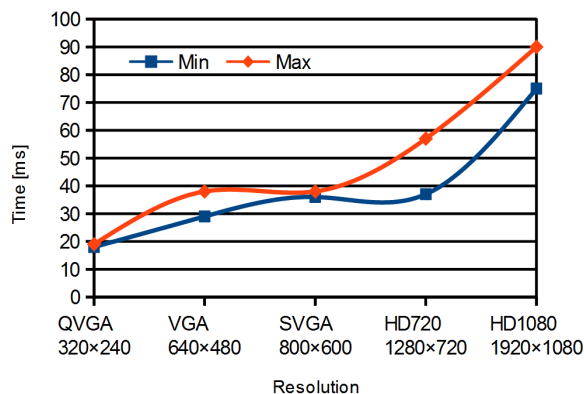


Figure 6: Comparison of execution time of the proposed system.

CPU Intel Core i7 - 2,4 GHz, GPU NVIDIA Quadro K1000M video card and 8 GB RAM. In the experimental testing, both approaches (overlap and trajectory check) for vehicle counting had the same execution time. Tested application was compiled without optimization and with GPU support. Video importing was performed by Microsoft Direct Show framework.

From the acquired results it can be concluded that real time vehicle detection can be performed on SVGA and lower resolutions using a standard PC computer. On SVGA resolution, 37 [ms] is required to process a single frame. This enables maximum frame rate of 27 [fps]. At QVGA resolution, 52 [fps] can be achieved with 19 [ms] required to process a single frame. It can also be concluded that approach with trajectory check gives better results (accuracy) than approach with overlap check. In second testing application was fully optimized by compiler and FFMPEG framework was used for video importing. Achieved results in the second testing show that application with GPU support and capability of executing on 8 threads by CPU can achieve much faster execution. At SVGA resolution, execution time for each frame was 17 [ms] which enables processing of 58 [fps]. At QVGA resolution, execution time was 7 [ms] which gives capability of processing 142 [fps]. The highest resolution in which application can still perform real time image processing is HD720 resolution with frame rate of 35 [fps]. In Fig. 8, ratio

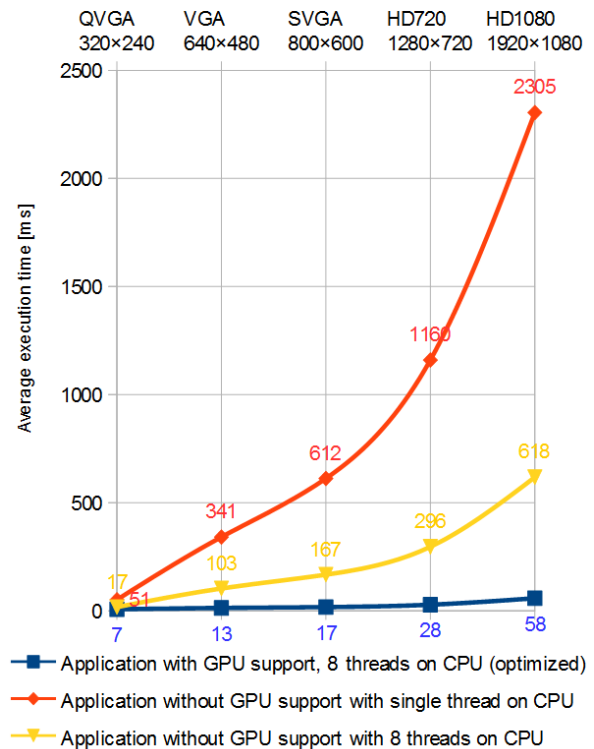


Figure 7: Comparison of application execution time with and without GPU and multi-thread capabilities.

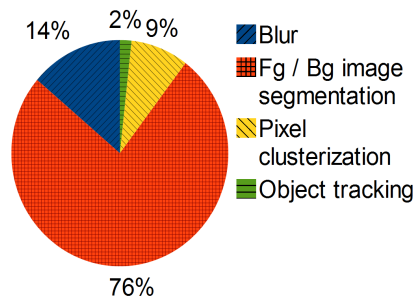


Figure 8: Execution time distribution between applied image processing tasks.

between execution time of a specific image processing task and overall execution time of the implemented application is given.

For the purpose of determining the efficiency of the proposed system with GPU and CPU multi-thread support, implemented application was modified to run without GPU support (only on CPU). At SVGA resolution application execution time for each frame was 167 [ms] (5 [fps]) with multi-thread capability (8 working threads) and 612 [ms] (1 [fps]) without multi-thread capability (single thread). In Fig. 7, comparison of previously mentioned version of implementations is given. All developed version of applications provide same results regarding vehicle detection accuracy (number of hits, FP and FN detections).

5 CONCLUSION AND FUTURE WORK

In this paper a system for vehicle detection and tracking on multiple lanes based on computer vision is proposed. Developed system uses only one camera to detect and track vehicles on a road with multiple lanes. First testing results are promising with vehicle detection accuracy of over 95%. Methods used in the proposed system are easy to implement and to parallelize. They are also suitable for executing in GPU and CPU multi-thread environments enabling real-time capabilities of the proposed system. Implemented vehicle trajectory tracking currently does not use any vehicle dynamics and predicts the tracked vehicle pose for one succeeding frame only.

From the execution time distribution analysis results can be concluded that algorithm for Fg/Bg image segmentation is the most slowest part of the application and it consumes 76% of the total application execution time. Further optimization of this algorithm would lower system requirements of the application and increase maximum resolution at which real time image processing can be achieved. It would also allow other complex algorithms (vehicle classification, license plate recognition, etc.) to be implemented into the system and executed in real time.

Future work consists of developing a multiple object tracking system which would estimate vehicle trajectory based on a vehicle model with dynamics included. Additionally, it is planned to develop a system which will perform vehicle classification and therefore separate vehicles by their type. This will enable detection and tracking of vehicles that are coming to standstill on crossroads.

6 ACKNOWLEDGMENTS

This work has been supported by the IPA2007/HR/16IPO/001-040514 project “VISTA - Computer Vision Innovations for Safe Traffic” which is co-financed by the European Union from the European Regional and Development Fund and by the EU COST action TU1102 - “Towards Autonomic Road Transport Support Systems”. Authors wish to thank Marko Ševrović and Marko Šoštarić for obtaining the road traffic video, and to Sanja Palajsa for preparing the test video.

7 REFERENCES

- [Con14] Pirc, J., Gostiša, B. Comparison between license plate matching and bluetooth signature re-identification systems for travel time estimation on highways, in Conf. proc. ISEP, 2014.
- [Con06] Bai, H., and Wu, J., and Liu, C. Motion and haar-like features based vehicle detection, in Conf. proc. MMM, 2006.
- [Con12a] Braut, V., and Čuljak, M., and Vukotić, V., and Šegvić, S., and Ševrović, M., and Gold, H. Estimating OD matrices at intersections in airborne video - a pilot study, in Conf. proc. MIPRO, pp.977-982, 2012.
- [Con13] Kovačić, K., and Ivanjko, E., and Gold, H. Computer vision systems in road vehicles: a review, in Conf. proc. CCVW, pp. 25-30, 2013.
- [Pon07] Ponsa, D., and López, A. Vehicle trajectory estimation based on monocular vision, Pattern recognition and image analysis, Vol. 4477, pp. 587-594, 2007.
- [Con00] Stanisavljević, V., and Kalafatić, Z., and Ribarić, S. Optical flow estimation over extended image sequence, in Conf. proc. MELECON, Vol. 2, pp. 546-549, 2000.
- [Con07] Tanaka, T., and Shimada, A., and Arita, D., and Taniguchi, R.I. A fast algorithm for adaptive background model construction using parzen density estimation, in Conf. proc. AVSS, pp. 528-533, 2007.
- [Con12b] Xiaoli, H., and Dianfu, H., and Xing, Y. Vehicle window detection based on line features, in Conf. proc. IMSNA, Vol. 1, pp. 261-263, 2012.

VisNow – a Modular, Extensible Visual Analysis Platform

Krzysztof S. Nowinski
University of Warsaw, ICM
Prosta 69
00-838 Warsaw, Poland
know@icm.edu.pl

Bartosz Borucki
University of Warsaw, ICM
Prosta 69
00-838 Warsaw, Poland
babor@icm.edu.pl

ABSTRACT

A new, dataflow driven, modular visual data analysis platform with extensive data processing and visualization capabilities is presented. VisNow is written in Java, easily extendable to incorporate new modules and module libraries. Dataflow networks built with the help of interactive network editor can be wrapped into stand-alone application for the end users.

Keywords

Scientific visualization, modular systems, dataflow driven system, medical imaging

1. INTRODUCTION

The VisNow system is based on longtime experience of usage and development of AVS systems (Advanced Visual Systems Inc., AVS 3-5 and AVS Express), IBM Data Explorer and several other general visualization systems [Peik07] [Hans04] [Para10][HPV13]. VisNow implements the concept of dataflow driven, modular system. The user builds a network of modules reading, processing and mapping data. Generic data structures are passed from output ports to input ports and the data processing is controlled by the user-manipulated parameters.

2. MODULAR STRUCTURE

All functionality of VisNow, including data input/output, logical and numerical processing (a.k.a. filtering), mapping of numerical data into geometric object and rendering these objects, are implemented as modules. Each module is capable of processing of data according to current parameter setting and outputting the results. In the case of input modules processing data means usually reading them from local disk or from some remote source. VisNow modules process two types of data: a **Field** encapsulating a discrete representation of data defined over an 1-, 2- or 3-dimensional area and a **GeometryObject** encapsulating a Java3D geometry. It should be noted that the distinction between filtering and mapping modules is somewhat blurred: filtering modules output simultaneously graphical rendering of the results to be shown in the viewer window and mappers usually output the results as

data object. For example, the module interpolating data to a regular mesh outputs the graphic rendering of the resulting field and the isosurface module outputs both a geometry object and an irregular mesh with interpolated data. Thus, the GUI of a module consists usually from **Computation UI** controlling module parameters, e.g. axis and location of a slice or input file name, and **Presentation UI** controlling color mapping of data and details of the presentation of points, lines, surfaces and volumes.

2.1. VisNow Granularity

The majority of existing systems based on the dataflow paradigm implement a fine-grained concept of relatively small building blocks. This model requires building of complicated networks to perform even simple tasks. In particular SciRun, IBM DX and other systems, require to instantiate “technical” modules like a colormap manager, 3D scene etc. to make the first geometrical object visible.

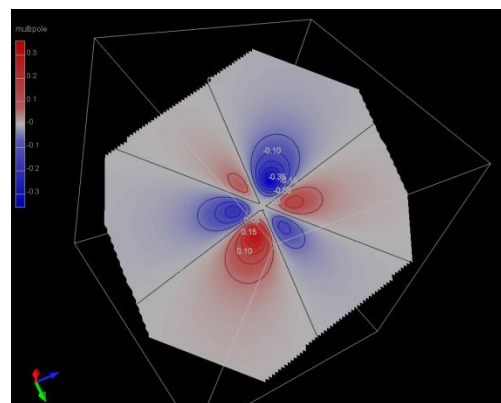


Figure 1. An example of VisNow visualization

In contrast, the VisNow system uses simple networks of high level modules: it is enough to use a reader, a slice module, an isolines module and the build-in viewer to obtain a reasonable visualization of a 3D data set – an application of read-in and see principle.

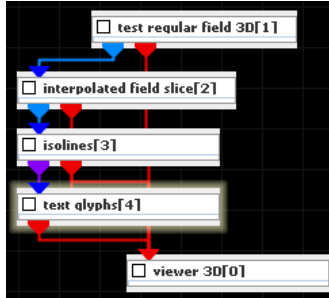


Figure 2. VisNow network producing Fig.1

VisNow provides fairly precise estimation of default parameter values. In particular, the mapping modules that could create extremely large geometries automatically downsize the input to reasonable dimensions leaving the final control over the geometry size to the user. As an example, glyph visualization of a vector field over an 512x512x512 mesh will be automatically downsampled to 64x64x64 mesh.

To simplify the interaction of the user with the network and module controls network area, module controls and the viewer display are synchronized. The user can pick a geometry object in the 3D window highlighting in the network area the module that created this object and bringing up the module controls.

3. VISNOW DATA STRUCTURES

VisNow uses internally a single, universal, abstract **Field** data type describing a discretization of an area in the Euclidean 1-, 2- or 3-space together with a set of numeric and non-numeric data defined over this area. The Field data type has two implementations – a **Regular Field** type and an **Irregular Field** type. Basically, a field consists of three basic components – the obligatory **structure**, the (explicitly provided or implicitly defined) **geometry** and a (possibly empty) set of **values**.

3.1. Regular Field

The Regular Field type covers all data sets with a **regular structure**, that is, a one- two- or three-dimensional table structure that is characterized by a simple list of **dimensions** – e.g. a 100,000,000-long time series, an 1920x1080 HD image or a 512x512x300 CT scan. In the simplest case, the geometry is defined implicitly with the node p_{ijk} located at the point (i,j,k) . In the general case of a regular **curvilinear field** all coordinates of all points can be provided explicitly, and in an intermediate

setting the user can set **affine coordinates** consisting of the origin point p and one, two or three cell vectors v_0, v_1, v_2 .

3.2. Irregular Field

The Irregular Field type requires explicit definitions of both structure and geometry. The structure is determined by the number of nodes and a list of **Cell Sets**. Each cell set is a collection of cells (point cells, segments, triangles, quadrangles, tetrahedral, pyramids, prisms and hexahedra). This data type covers basically all data sets occurring in FEM structural and CFD computations, but can be (and is) used for visualization of molecular structures, abstract graph presentation etc. Currently, VisNow does not support arbitrary polygons/polyhedral requiring off-line triangulation of such sets.

3.3. VisNow Data Values

VisNow uses a generic **Data Array** type to hold a variety of simple numeric (unsigned byte, short, int, float and double), complex, logical, string and generic Java object data. A data array holds the proper (flat) array of values of corresponding type and some additional data, e.g. name, physical unit, minimal and maximal values etc. Each data item can be a scalar, a vector or a (possibly symmetric) matrix. Thus, vector or tensor data can be processed natively with VisNow.

3.4. Time Dependent Data

The data values and node coordinates can be static or dynamic (defined for a list of **time moments**). It is sometimes important to have different data defined for different lists of time moments (e.g. numeric forecast outputs provide atmospheric pressure and ground level temperature for each five-minute time step while precipitation data are integrated over one hour intervals and orography or land cover are definitely static).

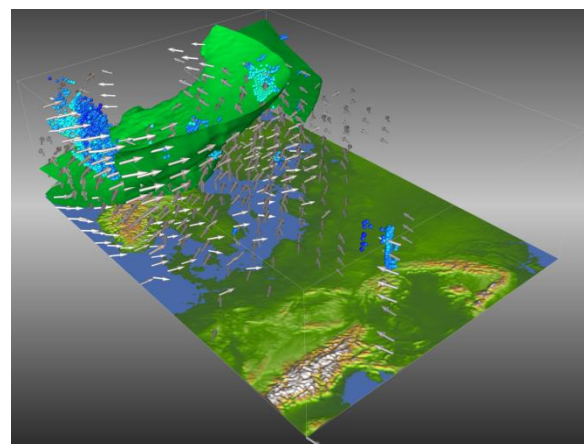


Figure 3. Numeric weather forecast data in the (x,y,t) coordinates: orography shown at the base, a pressure isosurface shows emergence of low pressure area,

wind and rain shown as glyphs

VisNow allows different lists of time moments for each variable (data array or node coordinates) providing a reasonable, piecewise linear interpolation for a given time moment. The data can be dynamically modified with time steps added or removed, and VisNow takes care for consistent interpolation. In the case of 2D fields, the time dependent data can be converted to a stack of slices of a 3D field – see Figure 3.

The only limitation of the VisNow time dependent data capabilities is the requirement of constant data structure (topology).

4. STANDARD VISNOW MODULES

4.1. Input/Output

VisNow supports a basic set of data formats with all types of images, simplified raw volume, AVS field and own VisNow regular field metafile format. The last file format is designed as a universal metafile allowing to read in ASCII or binary files containing static or time dependent coordinates and values of a regular field. In addition, a reader capable of browsing DICOM data is available for medical imaging data input, and an interactive trial-and-error method of ingestion of volumetric data of unknown dimensions is provided.

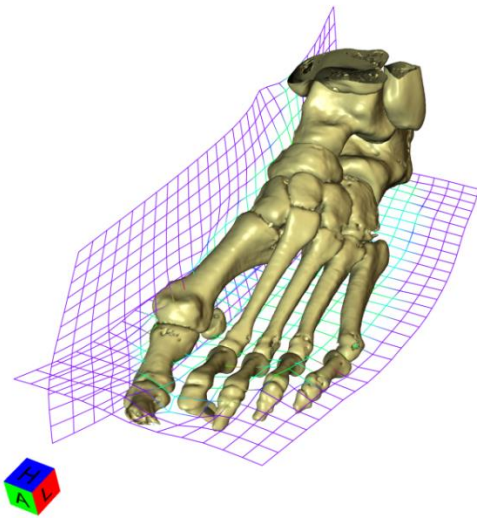


Figure 4. CT data read in from a DICOM file and deformed to match a patient image.

Unfortunately, the irregular data are so complicated that there is no way of developing a metafile that could describe all particular file formats. Thus, VisNow offers a set of modules capable of reading in AVS UCD, Ansys Fluent, EnSight and VTK data, usually with some limitations. Every partial result of

VisNow data processing can be stored in the VisNow (ir)regular field format.

VisNow can access data either by a local disk path or from remote URL. In addition, VisNow supports browsing of UNICORE grid resources either with the use of a grid virtual filesystem browser or with a grid bean for inserting a GridFTP data transfer as part of an UNICORE workflow.

The content of the viewer window can be written in any of the Java supported image formats in arbitrary resolution (not limited to the screen resolution). In addition, an animation can be stored and converted to the MPEG-4 movie format.

4.2. Data Modification (Filtering)

VisNow provides standard operations (downsizing and cropping regular data, interpolation to user defined meshes, simple arithmetic on data etc.). Advanced numerical processing modules include differential operations on regular fields with arbitrary geometry, FastFourier transform, convolution with user defined and editable kernel etc. Data calculations can be performed by simple mathematical formulas.

Advanced image denoising, segmentation and skeletonization algorithms have been implemented as elaborated data filtering modules. These modules are mainly used in medical applications.

4.3. Mapping of Data to Geometries

Each type of VisNow data (IrregularField, 1D-, 2D- and 3D-RegularField) has a default visualization mode. VisNow provides elaborated methods of mapping component values to colors including easily modifiable continuous or quantized colormap.

In the case of more than one data component it is possible to map one component to hue and modify the brightness or saturation by another component, e.g. mapping electrostatic potential to hue and field intensity to brightness. Blending of grayscale mapped anatomical data with physiological data mapped with a rainbow datamap is also a standard VisNow feature. In addition, the object transparency can be controlled by yet another selected data component.

The standard mappers library includes slicing, volume rendering and isosurfacing of 3D data, graphing and isolines creation in the case of 2D data and simple graphing of 1D data. Glyphs and text glyphs can be used for the representation of data at selected nodes. Streamlines, animated streamlines and object flow animations are available in the case of vector data components.

VisNow provides extensive capabilities of annotating the generated images. Colormap legends and coordinate axes are available with much attention paid to the clean and flexible labelling. The user can label field values by text glyphs, 3D annotations can

be located in arbitrary points of the 3D scene and 2D annotations (titles) can be displayed .

4.4. Viewers

VisNow provides a specialized 2D viewer and a simple graph viewer in addition to the main 3D viewer. A configurable 3D field viewer providing volume rendering and/or a set of orthogonal slice views helps to visualize 3D medical imaging data and an orthogonal viewer generates engineering type presentations.

The 3D viewer is the basic visual interaction node with the standard picking, geometric modification and (to the limited extent) object drawing.

5. IMPLEMENTATION AND EXTENDABILITY

VisNow is implemented in Java with Java3D as its graphic interface to GL and is structured as a set of NetBeans projects.

Each VisNow module is encapsulated in a Java package holding its main class, an XML description file, and (usually) classes holding module parameters and a GUI panel. In addition a set of XML library descriptions allow to choose between basic, standard and enhanced module libraries.

5.1. Extending VisNow

New module libraries (plugins) can be created as separate NetBeans projects importing standard VisNow JAR files and following the package structure described above. New libraries can be dynamically added and modified at the runtime. Fast Java compilation opens an interesting possibility of usage of VisNow as a sort of an integrated development environment. The user can encapsulate newly implemented algorithm into the VisNow module and use sophisticated GUI for program parameters and visual debugging. As an example, the development of non-rigid 3D registration of CT data (see fig. 4) has been done entirely within the VisNow system with many algorithmic and implementational problems found and resolved under constant visual control.

Any module network created in VisNow can be easily converted to a Java main class and released as a stand-alone application without the network GUI. Such form of the application suits well the needs of an end user.

5.2. Java Specific Problems

Java, together with NetBeans provides excellent environment for the development of large, complicated projects. Nevertheless, some basic limitations should be taken into account or overcome. The object paradigm of the language with significant memory overhead and the Java3D data access forced to use flattened data arrays. For example, coordinates of a field of N nodes are processed as a single array of the length $3*N$. On the other hand, the current hardware developments and increasing data size displayed a very serious Java language limitation. Array indices and collection sizes are 32-bit integers as. In effect, all Java data structures are limited to the size of at most 2^{31} items.

We are currently in progress of overcoming this limitation by the use of a library of non-standard array-like data structures developed in ICM. The JLargeArrays library allows to declare and use arrays indexed by long integers and thus the computational capabilities are limited only by physical memory size. The basic data structures and several basic data reading, filtering and visualization modules are already converted to JLargeArrays and we expect the next release of VisNow at least partly capable of large data processing.

6. VisNow AVAILABILITY

VisNow is available under GPL Classpath Exception public license with the binary installers for Linux, Windows and Mac OS, accessible from <http://visnow.icm.edu.pl>. The sources of VisNow and JLargeArrays are available at the GitHub repository.

7. REFERENCES

- [Peik07] Peikert R. Visualization systems. SciVis 2007, http://graphics.ethz.ch/teaching/former/scivis_07/Notes/Handouts/11-visSystems.pdf
- [Hans04] The Visualization Handbook, ed. by C.D. Hansen, C.R.Johnson, Elsevier 2004
- [Para10] ParaView guide, Kitware, Inc. Version 4 (August 2012)
- [HPV13] High Performance Visualization, ed. by E. Wes Bethel et al., CRC Press 2013

A new way of Rich Image Representation (VectorPixels)

Alain Simons, Edmond Prakash and James Wood

University of Bedfordshire

University Square

UK LU1 3JU, Luton,

United Kingdom

Alain.Simons@beds.ac.uk | Edmond.Prakash@beds.ac.uk | Jim.Wood@beds.ac.uk

ABSTRACT

For decades the same technologies has been used for generating images on computer systems. On the one hand you have bitmap and on the other hand vector technology. Bitmap or pixel technology is mostly used for the representation of rich colour images. The other available imaging technology, vectors, is mainly used to present logos and drawings without photo realistic details.

But both techniques have their disadvantages and advantages. In recent years a lot of research has been done to combine the advantages of both techniques in a comprehensive solution. Much research focused on giving a vector illustration a realistic picture look.

There was no development from scratch and that is exactly what this paper wants to propose, an algorithm to invent the pixel again. By always applying new developments on existing technologies generating images became a mix of various techniques and a rather complex matter.

With the use of VectorPixels (VP) the concept of a pixel will be redefined. In contrast with a pixel (Picture Element) a VectorPixel will be defined by a mathematical description and be resolution independent.

Keywords

Pixel, VectorPixel, image description, resolution independent, enlargement.

1. INTRODUCTION

Every day people are looking at images on a personal computer, tablet or mobile device. Programs used on these computer systems are becoming more complex and the use of images most specific on websites has increased significantly in recent years. Instead of working on a wired network, computer system users are using more often wireless networks. Wireless networks are becoming faster but a reduction of data transmission enable a more fluent browsing. VectorPixels would also minimize this problem by

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

using a reduced file size. This can be done by using

Scalable Vector Graphics (.svg) to store traditionally files. The .svg file format is supported by a lot of vector based software but delivers a large file when saving rich images.

A second issue is that making enlargements on

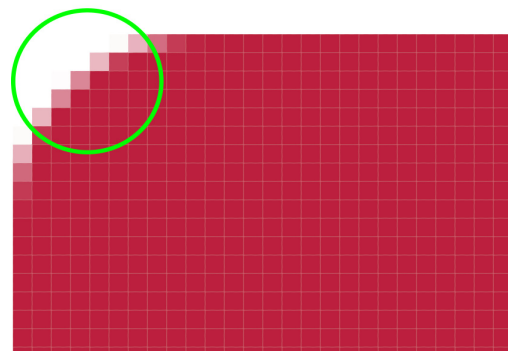


Figure 1. Jagged corner when using a bitmap solution when enlarging.

bitmap images will always end pixelated ((Figure 1).

Lines for example will be jagged and a mixture of pixels with different values of shades will be generated. This technique works fine when an image is viewed at 1:1 ratio and not enlarged. VectorPixels aims to provide a solution for this problem and

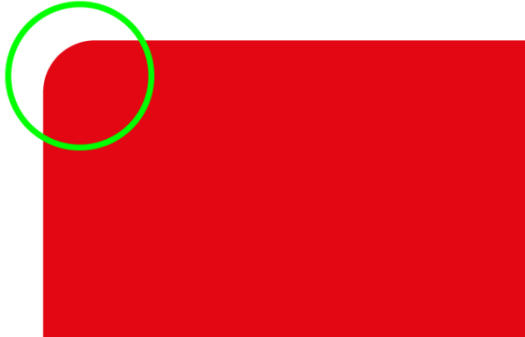


Figure 2. Smooth vector based corner when enlarged.

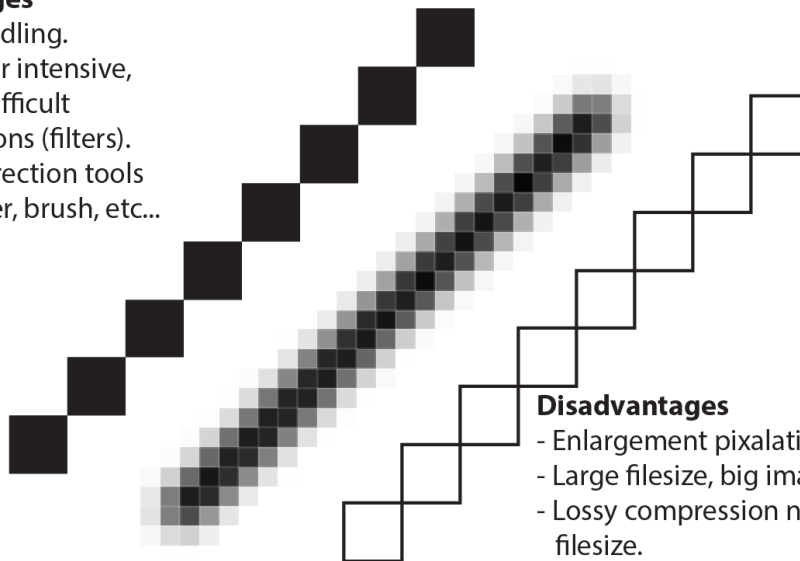
achieve the same image quality as obtained with vectors (Figure 2) with a “Richer Image”. Vectors present a sharply defined shape but there is no “natural look” or smooth color transition. The new proposed approach wants to keep the advantages of both imaging technologies but get rid of their disadvantages in particular, lack of smoothness for vector-based solutions and pixelated enlargements for bitmaps (Figure 3, 4).

2. RECENT SOLUTIONS

Research done by Orzan, Bousseau et al. [1] describe a new vector based tracing method for rendering smooth gradients from diffusion curves. This technique improves the smoothness of transition

Advantages

- Easy handling.
- Processor intensive, but no difficult calculations (filters).
- Easy correction tools like eraser, brush, etc...



Disadvantages

- Enlargement pixelation.
- Large filesize, big images.
- Lossy compression needed for smaller filesize.
- Large compression gives quality loss.

Figure 3. Advantages and disadvantages of using pixels to build up Rich Images.

between shapes but not the overall quality of the image. Also in this case, image information is added that wasn't there before. This research is partly based on a paper published by Elder and Goldberg [3] in March 2001 which used edge maps for certain image adjustments.

Tracing is an important issue in creating VectorPixels. At a first stage there is a simple line tracing, secondly an Edge tracing is carried out. Edge tracing has been the subject of a lot of research. There are different kinds of Edge detection like Ant-Based Detection by Aydin [9] which uses a state transition function based on 5x5 edge structures. Sobel is another possibility a technique used by Kunal [10] which uses an edge mask generated by a fast edge detector. Generating VectorPixels is complex, given that different techniques have to be combined to get a result. These techniques are line tracing, edge tracing and generating the VectorPixel itself. Also a VectorPixel has to be positioned with shape definition, fill, etc... Currently there is no solution that combines all these techniques with each other.

3. PROPOSED SOLUTION

Our solution is a new algorithm called VectorPixels and differs much from solutions proposed in the past by other researchers. Despite this it uses existing technologies. VectorPixel starts from a point that is defined by vectors and there is no such thing as a resolution grid. Instead of using a bitmap grid to position pixels, VectorPixels are positioned relative to a reference point. This reference point will be in the left upper corner of the image. A logarithm set up for VectorPixels can be seen in figure 5.

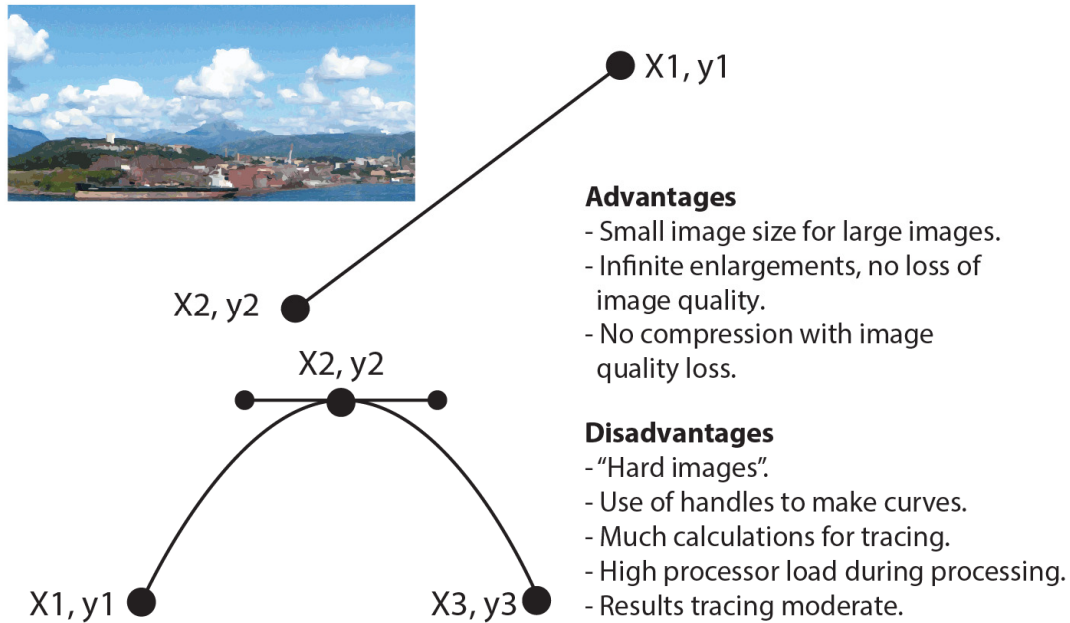


Figure 4. Advantages and disadvantages of using vectors to build up Rich Images.

This algorithm consists of five parts:

- Tracing mode
- Converting pixels to VectorPixels.
- Retracing edges.
- Smoothing.
- Export.

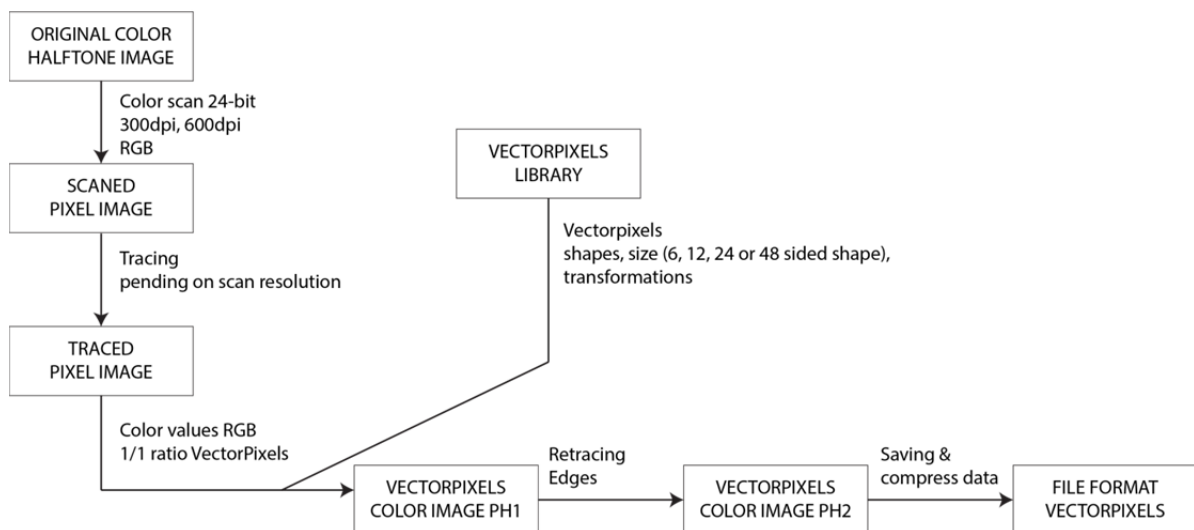
Because there are no capturing devices available to output VectorPixels based images, tracing of existing images is necessary. For testing purposes a simple line tracing is used. At a later stage other tracing techniques could be used or tested. Tracing will detect edges in the image by difference in color values of pixels. These edges can be rebuilt to be resolution independent with VectorPixels. Note a main difference with existing solutions using a closed

path to represent an area of similar pixels (Figure 6). With VectorPixels the area is built up with similar "vectorized" pixels. The shape and position of a VectorPixel can also vary depending on the detail needed in an image when enlarged. Changing resolution is already embedded in VectorPixels when they are created. VectorPixels can also include a gradation in order to give softness to edges.

A library can be provided where different sizes or shapes of VectorPixels can be archived and used as references. This can offer great advantages when saving VectorPixels based images especially with compression. In this case only the position of a VectorPixel and a reference, need to be saved. Of course this is together with the properties of the specific VectorPixel such as Fill Color or gradient

ALGORITHM VECTORPIXELS

Version 1.0



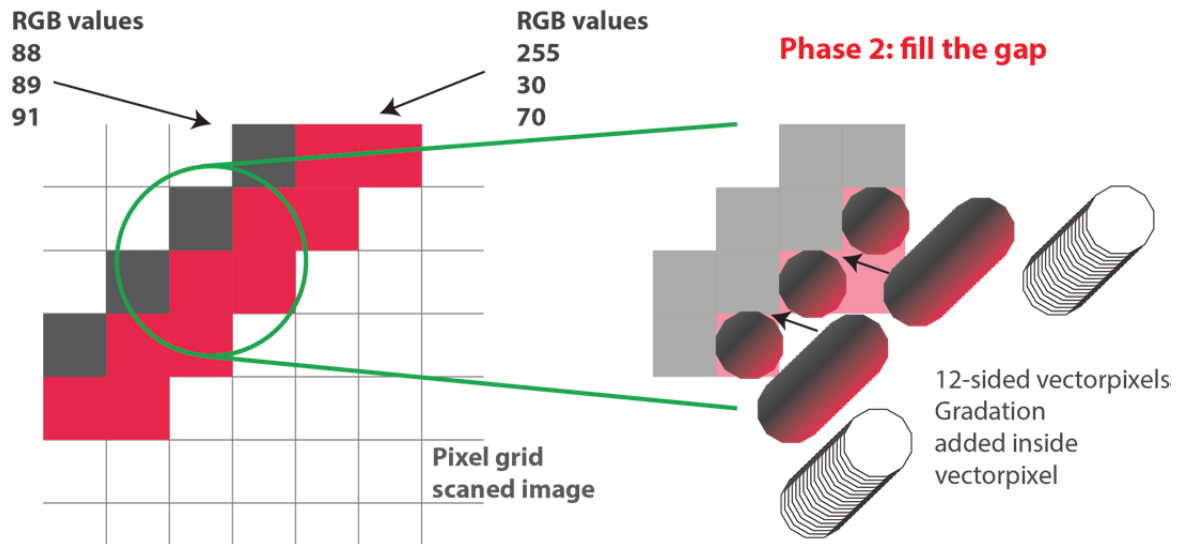


Figure 6. Tracing and smoothing VectorPixels.

and transformations. Making use of a gradient gives the possibility to make a smooth looking edge without additional programming. Because it is not working with a grid as with bitmaps, it is possible to position VectorPixels anywhere on the screen. For good order a maximum area (width and height) needs to be defined. Instead of using a fixed size in pixels a percentage could be used. VectorPixels can also overlap each other when a smooth curved line needs to be established (Figure 6). An algorithm (Figure 5) has been setup for first testing this new technique. Secondly this algorithm has to be realised in a programming environment which will have to deliver evidence that VectorPixels work effectively.

4. RESULTS

Our new developed algorithm is a theoretical one for the moment. Experimental research has now been carried out but useful results will be available by summer 2014. At a first stage OpenGL programming is being carried out to generate Algorithms to implement the VP technology on primitive shapes. This can be transferred to more complex cases such as Rich Images. In a third stage, a more practical application on Rich Images is provided with the use of VP in combination with Augmented Reality. Most of the augmented information is presented on top of an image (Figure 7). Information can be very divers, like distance between two points presented by a line

(and value) or a 3D representation of an object. AR technology is often a collaboration between software and hardware (glasses) for displaying additional



Figure 7. Augmented reality on top of a bitmap image.

information. Earlier research indicated that especially the lack of decent representation of extra information is a barrier for most users to get involved with Augmented Reality. In general there are also great expectation in the area of smoothing edges to be achieved more easily (Figure 8) than it happens with current technologies. Also reducing file size is an important goal. Another area where VectorPixels will be tested after been implemented in a 2D environment will be the use in a 3D environment [4]. Experiments will be conducted to measure the



Figure 8. VectorPixels shape define quality. Gradation included for smooth transition.

elapsed time for computation of VectorPixel algorithm. This could become one of the biggest drawbacks of VectorPixels due to the large processing requirements of calculations.

5. CONCLUSIONS

A lot of experimental research has to be done yet. Once first results become available a preliminary conclusion can be made. A schedule has been setup to get the first results published by the summer of 2014. The proposed algorithm is the baseline and could be refined when testing VectorPixels. This research is part of a larger whole. VectorPixels were first started to be used in 3D texturing. When VectorPixels are finally defined and working for 2D rich color images it can be transferred for use in 3D texturing. Especially in the area of shading it could be a great addition when Phong or Lambert shading needs to be calculated. This work will also benefit from the antialiased voxelization described in Prakash [11], [12].

6. REFERENCES

- [1] A. Orzan, A. Bousseau, P. Barla, et al. "Diffusion Curves: A Vector Representation for Smooth-Shaded Images", New York, ACM 56 (7), pp.101-108, July 2013.
- [2] J. Canny. "A Computational Approach to Edge Detection", MIT, IEEE Transactions on Pattern and Machine Intelligence PAMI-8 (6), pp. 679-696, November 1986.
- [3] J.H. Elder and R. Goldberg. "Image Editing in the Contour Domain", Canada, York University, IEEE Transactions on Pattern and Machine Intelligence, 23 (3), pp. 291-294, September 2000.
- [4] J. Behr et al. "A Scalable Architecture for the HTML5/X3D Integration Model X3DOM", *Web3D'10*, Los Angeles CA, 2010, pp.185-194.
- [5] J.C. Mong and D.F. Brailsford. "Using SVG as the Rendering Model for Structured and Graphically Complex Web Material", Electronic Publishing Research Group, University of Nottingham, ACM symposium on document engineering, pp. 88-91 November 2003.
- [6] W3C. (2010). Scalable Vector Graphics. Available: <http://www.w3.org/Graphics/SVG/>. Last accessed 29th March 2014.
- [7] Adobe Developer Center. (2013). Scalable Vector Graphics. Available: <http://www.adobe.com/devnet/svg.html>. Last accessed 29th March 2014.
- [8] E. Kim, X. Huang and G. Tan, "Markup SVG-A Online Content-Aware Image Abstraction and Annotation Tool", IEEE Transactions on Multimedia 13 (5), pp. 993-1006, July 2011.
- [9] D. Aydin, "Transitions on Computational Collective Intelligence", Springer Lecture Notes in Computer Science. 6220, pp. 39-55, May 2010.
- [10] R. Kunal, "Unsupervised edge detection and noise detection from a single image". Elsevier Pattern Recognition 46 (8), pp. 2067-2077, August 2013.
- [11] C. E. Prakash and S. Manohar, "Error Measures and 3D Anti-aliasing of Voxel Data", in Proc. of the Pacific Graphics 95, Taejon, Korea, 1995, pp. 225-239.
- [12] C. E. Prakash et al., "Voxel-Based Modeling for Layered Manufacturing", Computer Graphics and Applications, IEEE, 15(6), pp. 42-47, November 1995.

Lane Marking Detection in Various Lighting Conditions using Robust Feature Extraction

Ho-Jin Jang

School of Computer Science
and Engineering , Kyungpook
National University

KS002 ,Daegu, Republic of
Korea,

hobbangc@vision.knu.ac.kr

Seung-Hae Baek

School of Computer Science
and Engineering , Kyungpook
National University

KS002 ,Daegu, Republic of
Korea,

eardrops@vision.knu.ac.kr

Soon-Yong Park

School of Computer Science
and Engineering , Kyungpook
National University

KS002 ,Daegu, Republic of
Korea,

sympark@knu.ac.kr

ABSTRACT

Robustly extracting the features of lane markings under different lighting and weather conditions such as shadows, glows, sunset and night is the a key technology of the lane departure warning system (LDWS). In this paper, we propose a robust lane marking feature extraction method. By using the characteristics of the lane marking to detect candidate areas. The final lane marking features are extracted by first finding the center points of the lane marking in the candidate area then these center point pixels are labeled according to the intensity similarity along the direction of the vanishing point. The performance of the proposed method is evaluated by experiment at results using real world lane data.

Keywords

LDWS (Lane Departure Warning System), Lane Marking Feature Extraction, Lane Detection, Image Processing, Computer Vision

1. Introduction

Recent automotive industry mainly concerns about the safety and convenience of the driver and much research work on systems such as ITS (Intelligent Transport System) and ADAS (Advanced Driver Assist System) have been conducted. One of the main key technologies of ITS and ADAS is the LDWS (Lane Departure Warning System), a system which detects information of the current driving lane and warns the driver when the vehicle begins to move out of its lane. This system has become more famous and many studies are underway as it is designed to minimize accidents caused by the unconsciousness of driver.

The Lane Marking Feature Extraction used in LDWS system is based on conventional methods such as color model method[Sun06a], [Lee09a], edge information method[Lin10a], [Macek04a], B-Snake method[Wang04a], stereo method[Yaylor96a], [Hattori00a], and perspective transformation method[Aly08a].

In the color model method; the lane marking is first detected by converting the RGB image into the HSI (hue, saturation, intensity) color model and then binarizing it by using saturation and intensity values. The Lane marking is detected well in good environment conditions but, the detection rate

becomes very low and poor in low light situations such as sunset or dawn and even when shadow is mixed.

In the edge information method, Canny or Sobel filters are used to detect the lane edges. The line is then extracted by identifying the straight line components using the Hough space. Unlike the color model method, this method manages to detect the line in different lighting conditions such as sunset or cloudy environments without being affected by the color temperature. But the extraction of the lane marking becomes more difficult if edges other than straight line components are detected.

In this paper, we generate an accumulative intensity difference image of the lane marking and road areas by using the characteristics of the lane marking. Then we obtain the center point candidate image of the lane marking by using the distribution kernel of the accumulative image and label the pixels of the image in the direction of the vanishing point. The current driving lane is then detected by the Hough transformation of the lane marking feature extraction image.

Section 2 of the paper describes the lane feature extraction process, Section 3 describes the procedure used to detect the lane marking using feature points, Sections 4 evaluate the recognition rate of the

proposed method through experiment results and Section 6 depicts conclusion.

2. Lane Marking Feature Extraction

In our paper, we have made three assumptions for robust lane marking feature extraction in different lighting conditions. Three assumptions are as follows.

- First, the lane marking has a constant width.
- Second, the intensity value of the lane marking area is higher than the intensity value of the road area.
- Third, the driving direction and the lane marking are parallel.

As above mentioned assumptions are invariant to light conditions; they are used to detect the candidate region of the line in LDA (Line Difference Accumulation). The center points of the candidate area are then detected in the process of LCC (Line Center Candidate) and the lane marking is detected in the LCC-R (Line Center Candidate-Refinement) process by removing the components except the detected center points. Fig. 2 shows the proposed lane marking detection procedure.

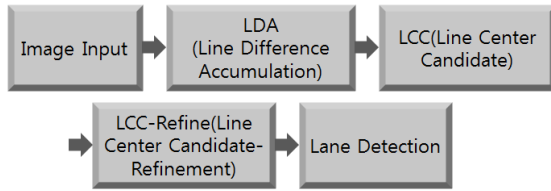


Figure 1. Flow diagram of the proposed algorithm

2.1 Line Difference Accumulation(LDA)

An input image is first converted into a gray image. The intensity difference of image $I(x, y)$ and $I(x \pm d, y)$ is calculated using Equation (1) where $I(x, y)$ is pixel intensity and d is an offset value of x .

$$\begin{aligned} \text{Val}(x, y) = & 2 \times I(x, y) - I(x + d, y) \\ & - I(x - d, y) \end{aligned} \quad (1)$$

After generating $D(x, y)$ according to Equation(2); Equation (3) is used to create $H(x, y)$ LDA image by increasing the value of d from d_{min} to d_{max} .

$$\begin{aligned} D(x, y) = & \begin{cases} 0, & \text{if } (I(x, y) < I(x + d, y) \text{ or } I(x, y) < I(x - d, y)) \\ \text{Val}(x, y), & \text{otherwise} \end{cases} \end{aligned} \quad (2)$$

$$\begin{aligned} H(x, y) = & \sum_{d=d_{min}}^{d_{max}} D(x + d, y) + \sum_{d=d_{min}}^{d_{max}} D(x - d, y) \end{aligned} \quad (3)$$

$$\begin{aligned} d_{min} &= \mu \times (y - V_y) \\ d_{max} &= \mu \times (y - V_y) + 3 \end{aligned} \quad (4)$$

In Equation (4), y is the y -axis coordinate used to calculate d and V_y is the y -axis coordinate of the vanishing point. Even though the width of the lane marking is said to be same; the width of the lane marking in image is proportional to the y -axis and as a result; d is changed according to the y -axis as it is expressed in Equation (4). μ is a weight factor of the lane marking width, and it is determined by camera calibration.

The intensity value of the road area in the gray level image is low and the intensity value of the lane marking is high. As the intensity difference between these two values is high; it allows us to acquire a potential candidate region. In the detected $H(x, y)$ image, one pixel can be a strong candidate of the lane marking if the intensity difference between the pixel and another pixel apart from d distance is high.

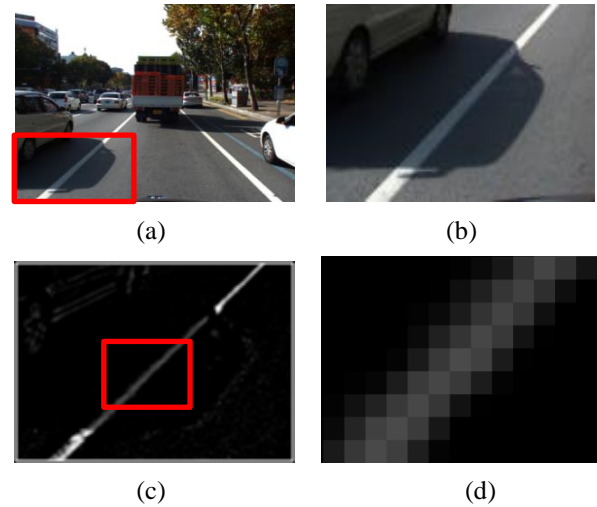


Figure 2. Result of LDA image. (a) input image. (b) cropped input image of the red box. (c) LDA image. (d) cropped LDA image of the red box.

Detected lane marking candidate area is shown in Figure.2 and the center point appears to have the highest intensity. As it contains the highest intensity, strong candidates can be detected without any difficulties even sunset or shadow areas contain lower intensity.

2.2 Line Center Candidate(LCC)

In this section we find the center of the candidate area using the LDA (Line Difference Accumulation) image which we have detected in the previous section. An LDA image pixel gets a higher intensity value as it becomes the center point of the lane marking and it gradually decreases toward the outskirts of the lane marking. The kernel which has a distribution function of $f(x) = 5 / (5 + x^2)$ (equation (5)) inside the search range is correlated and the largest pixel of the result is assigned as the center of the line.

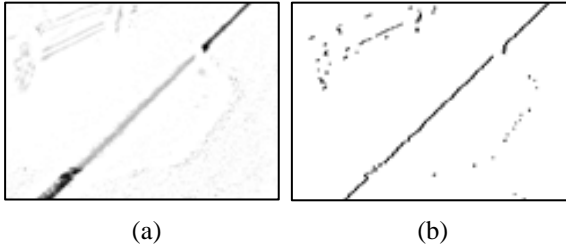


Figure 3. Result of LCC image. (intensity of the image was reversed) (a) LDA image. (b) LCC image..

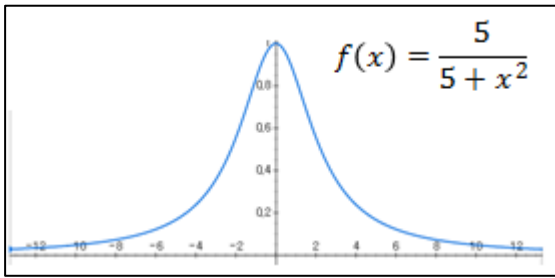


Figure 4. Distribution function, $f(x) = 5 / (5 + x^2)$

$$C(x,y) = \underset{i \in [-SearchRange, SearchRange]}{\operatorname{argmax}} \left(\sum_{j=-KernelSize}^{KernelSize} \left(\left(\frac{5}{5+j^2} \right) * H(x+i+j, y) \right) \right) \quad (5)$$

$$SearchRange = \alpha \times (y - V_y) + 3 \quad (6)$$

$$KernelSize = \beta \times (y - V_y) + 2 \quad (7)$$

α and β are arbitrary constants which are determined through experimentation. KernelSize determines the size of the distribution kernel whereas SearchRange determines the search range. We set the size of the kernel and the SearchRange to be proportional to the y-axis as the pixel width of the lane marking is proportional to the y-axis. According to Equation(5), the maximum value point of the distribution kernel in the search range is assigned as the lane center

candidate. Figure.3 shows the results of the Line Center Candidate.

2-3. Line Center Candidate Refinement (LCC-R)

The noise of the LCC image obtained in the previous section can be detected as a candidate region of the center of the lane marking.

In this section, we remove the noise of the candidate region by applying the *connected component labeling* (Figure5.) along the vanishing point direction starting from the bottom of the y-axis. In this process, if the number of connected pixels is less than five; connected area is not considered as a lane marking candidate and will be removed. Figure5 shows the labeling process. Figure.6 shows the Line Center Candidate Refinement results.

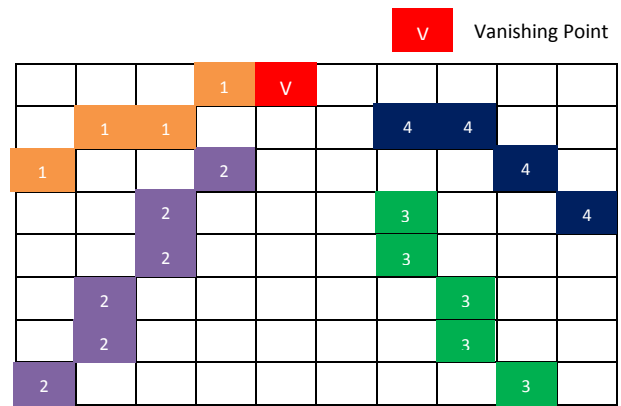
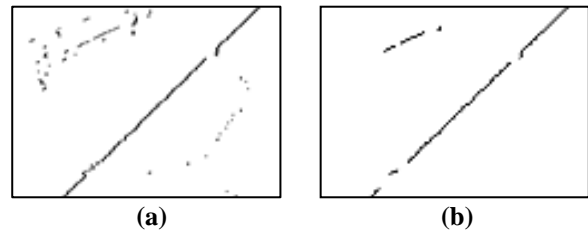


Figure 5. LCC-R image labeling process

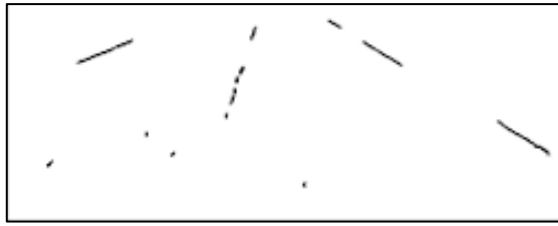


**Figure 6. Result of LCC-R image.
(a) LCC image. (b) LCC-R image..**

3. Lane Marking Detection

In this section, we detect the lane marking by using the LCC-R image which we obtained in the previous section.

In order to detect the lane marking, we apply the Hough transform [Yu97a] to the LCC-R image. The two Hough lines which have the highest intensity value of the LDA image in the vanishing point direction are determined as the last lane markings. Figure.7 shows the results of the final lane marking detection.



(a)



(b)

Figure 7. Results of lane marking detection.

(a) LCC-R image (b) lane marking detection image

4. Experiment

After mounting the Pointgrey's Grasshopper2 and a lens with focal length of 8mm in the car, we performed the experiment by obtaining image at 30 fps in a 640×480 resolution in a PC running Windows7 containing an Intel Corei5-3470 3.20GHz with a 8GB RAM.

If the Hough line is presented on the lane marking, it is assigned to the Success column. If a Hough line does not exist or presented on a non lane marking position; it is assigned to the Fail column. Figure 8 shows the results of lane feature extraction done in various lighting conditions.

	Total frame	Success (frame)	Fail (frame)	Recognition rate (%)
Daylight	1500	1493	7	99.53%
Sunset	1200	1184	16	98.66%
Shadow	800	756	44	94.50%
Night	1200	1189	11	99.08%
complex	1000	946	54	94.60%
Total recognition rate				96.27%

Table 1. Recognition in various environments

A higher overall recognition rate of 96.27% was obtained when we performed the experiment in complex environment conditions such as daylight, sunset, shadow and night.

Incorrect recognition occurs when guardrails or sidewalks are detected as they have the similar characteristics as the road lane marking.

If the paint of the line is tattered, it can be detected as noise and will be removed.

In cases of misrecognition; shadows of road side trees, buildings can be detected as the road line when the sunlight is projected parallel to the road in the same angle.

5. Conclusion

A robust lane marking feature extraction method in various light conditions was proposed. In this proposed method, the LDA (Line Difference Accumulation) image was obtained by using the independent characteristics of illumination and the LCC (Line Center Candidate) image was obtained by using the intensity distribution characteristics of the lane marking which was obtained by LDA. After removing the noise from the LCC image; the strong lane marking which was invariant to different environment conditions was obtained by detecting the LCC-R image. The processing speed also can be used for real-time detection at 22.31fps as the experimental results from a variety of lighting conditions gave a very high average recognition rate of 97.94%.

6. ACKNOWLEDGMENT

This research was supported by the MSIP(Ministry of Science, ICT & Future Planning), Korea, under the C-ITRC(Convergence Information Technology Research Center) support program (NIPA-2014-H0401-14-1004) supervised by the NIPA(National IT Industry Promotion Agency.) and the Industrial Strategic Technology Development Program of Ministry of Trade, Industry & Energy (MOTIE, Korea) [10040927, Driver-oriented vehicle augmented reality system based on head up display for the driving safety and convenience].

7. Reference

- [Sun06a] Tsung-Ying Sun ; Nat. Dong Hwa Univ., Hualien ; Shang-Jeng Tsai ; Chan, V.. Intelligent Transportation Systems Conference "HSI Color Model Based Lane-Marking Detection", Toronto, Canada, September 17-20, pp. 1168-1172, 2006.
- [Lee09a] Jin-Wook Lee ; Sch. of Internet-Media Eng., Korea Univ. of Technol. & Educ., Cheonan, South Korea ; Jae-Soo Cho , Computer Sciences and Convergence Information Technology, "Effective lane detection and tracking method using statistical modeling of color and lane edge-orientation" Dhaka, Bangladesh, Dec. 21-23, pp.1586-1591, 2009.
- [Lin10a] Qing Lin ; Dept. of Electron. Eng., Soongsil Univ., Seoul, South Korea ; Youngjoon Han ; Hernsoo Hahn , Computer Research and Development, 2010 Second International

- Conference on "Real-time lane departure detection based on extended edge-linking algorithm." Washington, DC, USA, May. 7-10, pp. 725-730, 2010.
- [Macek04a] Macek, Kristijan, et al. Proceedings of the IEEE/APS Conference on Mechatronics and Robotics. "A lane detection vision module for driver assistance." pp. 77-82, 2004.
- [Wang04a] Wang, Yue, Eam Khwang Teoh, and Dinggang Shen. Image and Vision computing 22.4 "Lane detection and tracking using B-Snake." 22.4: 269-280., 2004.
- [Taylor96a] C. J. Taylor, J. Malik, and J. Weber, Intelligent Vehicles Symposium, "A real-time approach to stereopsis and lane-finding," pp. 207-212, 1996.
- [Hattori00a] H. Hattori, Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE. "Stereo for 2D visual navigation," pp. 31-38, 2000
- [Aly08a] Aly, M. ; (2008). Comput. Vision Lab., California Inst. of Technol., Pasadena, CA "Real time Detection of Lane Markers in Urban Streets" pp. 7-12, 2008.
- [Yu97a] Yu, B., & Jain, A. K. Image Processing, 1997. Proceedings., International Conference on. Vol. 2 "Lane boundary detection using a multiresolution Hough transform" pp. 748-751, 1997.

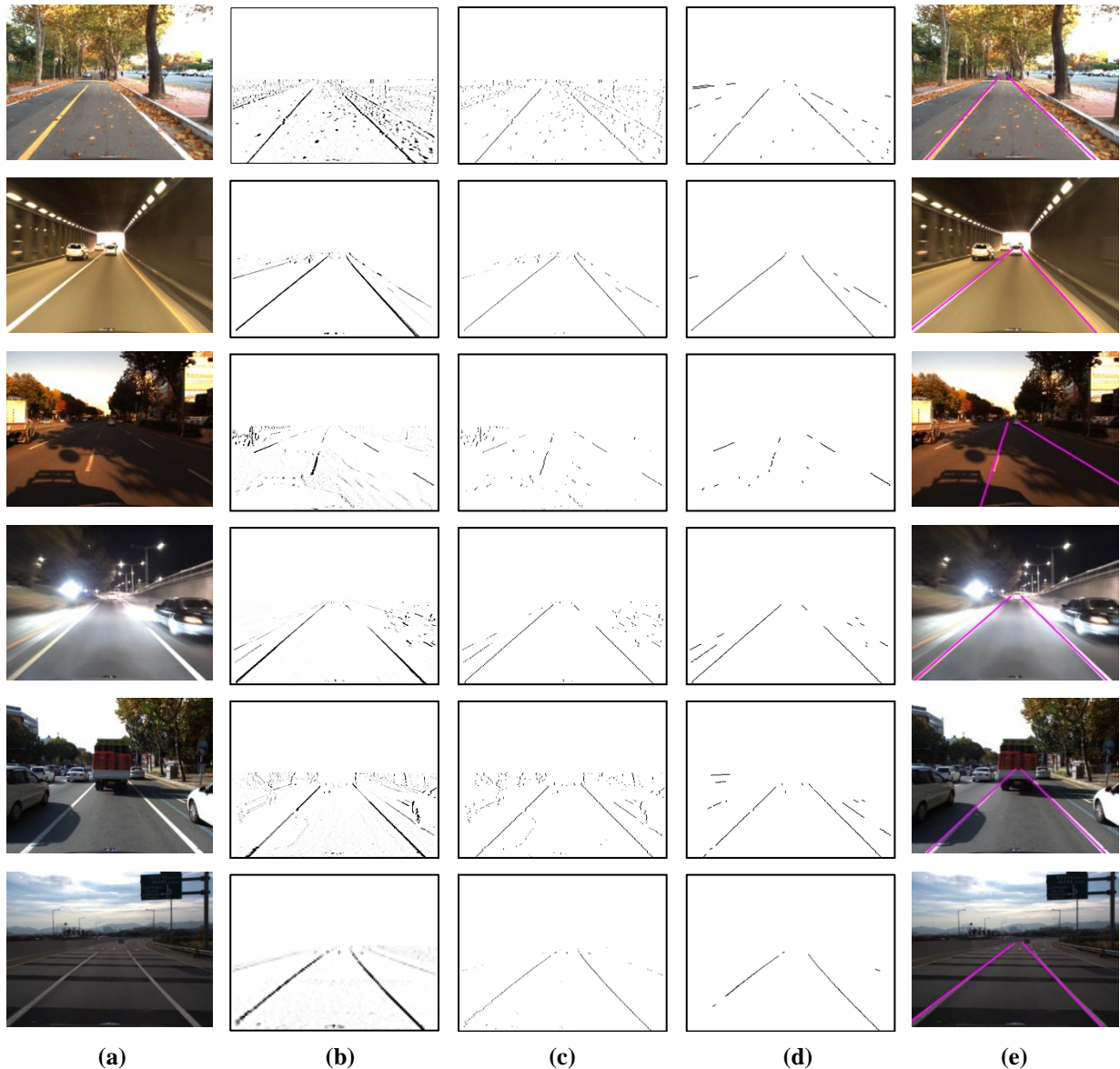


Figure 8. Result of lane feature extraction by various lighting conditions.
(a)input image (b)LDA image (c) LCC image (d)LCC-Refine image (e)lane marking detection image

Survey on Automated LEGO Assembly Construction

Jae Woo Kim
ETRI
218 Gajeong-ro
Yuseong-gu
Daejeon 305-700, Korea
jae_kim@etri.re.kr

Kyung Kyu Kang
ETRI
218 Gajeong-ro
Yuseong-gu
Daejeon 305-700, Korea
kangk2@etri.re.kr

Ji Hyoung Lee
ETRI
218 Gajeong-ro
Yuseong-gu
Daejeon 305-700, Korea
ijihyung@etri.re.kr

ABSTRACT

LEGO has been very popular toy in the world because it is attractive and fun to play with and stimulates one's creativity by providing means to conveniently assemble a variety of interesting shapes using the limited types of given bricks. However, it is hard for the beginners to design and assemble complex models they desire to make without instructions. Building a LEGO assembly manually usually requires a significant amount of trial-and-error. LEGO company therefore presented the LEGO construction problem in 1998 and in 2001. The problem statement is "Given any 3D body, how can it be built from LEGO bricks?" In this paper we will investigate the current research efforts to address the LEGO construction problem. We will review the problem definition, formulation, and a variety of approaches to solve the problem. We will discuss the data representations for input 3D polygonal models and the LEGO assembly structures, cost functions that will guide the search for the optimal solution, and various solution methods.

Keywords

Engineering, Design, LEGO, Construction, Optimization, Evolutionary, Algorithms

1. INTRODUCTION

In the 1940s in Denmark, LEGO was designed, developed, and produced by the LEGO company, the most successful toy manufacturer [Sma08, Sil09]. LEGO has been very popular toy in the world because it is attractive and fun to play with and stimulates one's creativity by providing means to conveniently assemble a variety of interesting shapes using the limited types of given bricks (Fig. 1) [Tes13, Ono13].

However, it is hard for the beginners to design and assemble complex models without instructions [Ono13]. Building a LEGO assembly manually requires a significant amount of trials-and-errors [Tes13]. LEGO company therefore presented the LEGO construction problem in 1998 and in 2001. The problem statement is "Given any 3D body, how can it be built from LEGO bricks [Tim98]?"

Researchers tried to develop softwares that can automatically generate LEGO assemblies and assembly instructions from the geometric specifications of the desired object. In most research efforts, 3D polygonal model data were used as specifications of the objects.

LEGO construction problem is simple and easy to understand. It is however hard to solve using mathematical or algorithmic approaches on computer because there exist a number of different ways to construct a LEGO model for an object specified by users [Sma08].

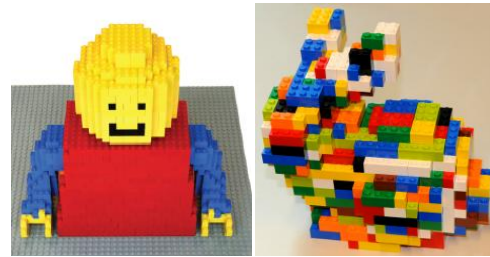


Figure 1. Computer generated LEGO representation(left) and real LEGO assembly(right) [Tes13]

Other than the applications for entertainment purposes described above, study on LEGO construction problem will have great practical value and contribute to other areas such as engineering design or engineering education because the process of LEGO assembly generation is similar to the generation of real engineering artifacts [Pey03]. Campbell et al. showed that how various physical and chemical principles related to nanoscale science and technology can be demonstrated using LEGO models [Cam12]; Wang et al. designed and developed a digital LEGO system that provide a generic representation of security protocols and used it in teaching students. The digital LEGO system helps the students conveniently understand these abstract concepts [Wan08]; Yip-Hoi and Newcomer used LEGO to teach CAD modeling techniques to engineers [Yip11].

Solution methods developed to solve the LEGO construction problem are related to other interesting

problems: Mitani et al. proposed a method to produce unfolded papercraft patterns of toy figures from 3D polygonal mesh data using strip-based approximation [Mit04]; Igarashi and Suzuki proposed a method to create close-fitting customized covers for three-dimensional objects [Iga11]; Xin et al. proposed a method to design and model burr puzzles from 3D geometric models [Xin11]; Lo et al. proposed a method to generate 3D Polyomino puzzle that constructs 3D surface model using Polyomino pieces [Lo09].

In this paper, we investigated a variety of approaches to solve the LEGO construction problem. In the next chapter, the problem definitions and formulations will be covered. In chapter 3, we will discuss the data representations for both of the input 3D polygonal models and the LEGO assembly structures. In chapter 4, we will introduce the cost functions that will guide the search for the optimal solution and in chapter 5 we will discuss various solution methods that have been used to solve the LEGO construction problem.

2. Automated LEGO Assembly Construction Problem

When a user has a design of an arbitrary object in mind, she will try to build an LEGO assembly with a significant amount of trials and errors. LEGO construction problem is to find the optimal way of converting 3D polygonal mesh data to LEGO representation given the number of LEGO bricks.

Smal defined the LEGO construction problem as the development of a software application that generates the LEGO building instructions for the given arbitrary real-world object [Sma08]. Here, the real world object is represented by 3D polygonal mesh models. The output of the LEGO construction software applications include LEGO model representations, 3D renderings of the LEGO model, and assembly instructions for building the LEGO models.

We need to simplify the problem by applying several restrictions that can reduce the search space for the solution [Gow98][Sma08][Tim98]. Real-world object that users desire to build must be firstly converted to an appropriate representation such as "legolised" representation. The legolised representation is a matrix whose elements can only have ones and zeros. The value one for an element denotes that the space is covered with a brick or a part of a brick, the value zero denotes an empty space [Sma08].

Bricks available in building a LEGO representation must be restricted to a limited number of types to save processing time for optimization by reducing the search space. Usually, "family" LEGO bricks and DUPLO bricks were used in previous research to

address the problem [Pet01]. To save time and reduce the number of bricks, the inside of the sculpture must be kept hollow as far as the connectivity and stability are kept. Colors can be ignored to save processing time because incorporating color information into the problem can increase another dimension of search space resulting in drastic increase in search space [Sil09].

There are two major performance criteria that the solution of the LEGO construction problem must satisfy. The first criterion is that the created LEGO sculpture must be connected and stable. Another criterion is that the conversion from an object model to a corresponding LEGO representation must be complete in a reasonable time period [Sma08].

The automated LEGO construction problem can be formulated as an optimization problem to find the optimal LEGO structure that best represent the input real-world object. In general, optimization techniques can be divided into two categories: one is a deterministic technique and the other is a stochastic technique. Deterministic approaches are used to find the globally optimal solution and they are appropriate to the problems whose solution space is relatively small. Efficient state space search methods such as branch-and-bound methods, or algebraic methods are usually used to find the globally optimal solution. When the solution space is extremely large and it is therefore not feasible to find global optimal solution, Stochastic technique can be used. Stochastic technique finds good solutions in a reasonable time period by using heuristics and probability theories that guides the search [Sma08].

The automated LEGO construction problem cannot be solved using deterministic optimization techniques because the solution space is extremely large. We therefore discuss stochastic optimization techniques to solve the LEGO construction problem [Sma08].

There are a variety of solution methods to address the optimization problems and greedy methods, local search, beam search, cellular automata and evolutionary algorithms were used to solve the LEGO optimization problem. From next chapter, we will discuss the approaches used to solve the automated LEGO construction problems.

3. DATA REPRESENTATIONS

The real-world object that the users desire to create are usually given as a 3D polygonal mesh models. Users can create the 3D mesh models using modeling softwares or can easily download them from the internet [Lam06]. The first step in solving the LEGO construction problem therefore is to convert a given 3D polygonal mesh models to a data representation that is appropriate for the process of LEGO assembly generation. A typical data representation for the real-

world object is a "legolised" model proposed by [Gow98].

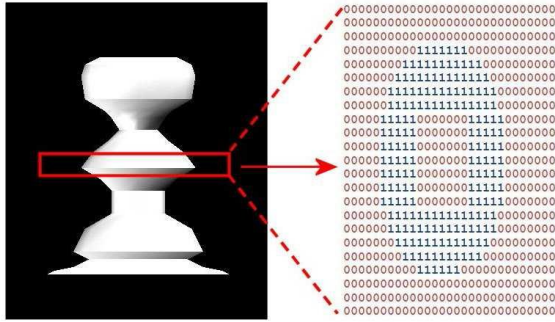


Figure 2. A 3D polygonal model and its legolised representation for a horizontal cut [Sam08]

Voxelization

Voxel representation is naturally employed by researchers to represent the real-world objects because basic LEGO bricks has the rectangular shapes that matches well with voxels. LamBrecht used ray casting technique in voxelizing the input 3D mesh models. Their approach casted a ray in a axis-aligned direction from the each column of voxels. The algorithm conducts voxelization by iterating through all the faces of the model in the cube and testing for intersection between the face and the ray [Lam06].

Silva et al. proposed a novel voxelization algorithm that uses point samples of the surface to determine which voxels each point belong to. Their algorithm assumes a uniform sampling of the surface. In general, it is however not guaranteed for most of the triangular meshes due to their shape irregularity with varying edge sizes. They therefore conducted a subdivision algorithm to transform the given mesh model into a uniformly sampled model where all lengths of all the edges are smaller than the user-specified resolution. They implemented their algorithms on the GPU to achieve the real-time performance [Sil09].

After or during the voxelization process current approaches hollowed the model to decrease the processing time by hollowing the model. This process is conducted by removing unnecessary bricks while stability is not damaged [Tes13][Lam06].

We have to consider the trade-off of the voxel resolution when we perform the voxelization process and specify appropriate resolution depending on the application purposes. If the resolution is high, we can represent the more detailed shapes of the model but it increases the processing time drastically. If the resolution is low, the voxelization algorithm runs fast but the quality of the model is not convincing. [Tes13]

LEGO Model Representation



Figure 3. Examples of basic LEGO bricks 1x1, 1x2, 1x3, 1x4, 1x6, 1x8, 2x2, 2x3, 2x4, 2x6, 2x8 [Ono13]

The simplest way of representing LEGO model is using the voxel representation. In this approach, each voxel can be identical to the unit LEGO brick of the size 1x1 or a part of a larger brick (Fig. 3) [Ono13]. In this approach, the voxel representation is converted to the LEGO representation within the voxel space.

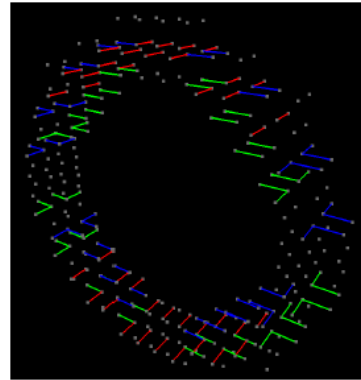


Figure 4. A legograph that represent the connectivity among the bricks in voxel space

In the beginning of the process, each voxel that the object covers is occupied by an unit brick and then replaced later by the larger bricks by merging the unit bricks considering connectivity among the bricks [Ono13].

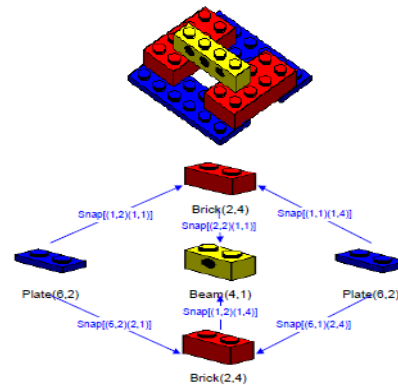


Figure 5. An example of a LEGO assembly structure and a corresponding assembly graph [Pey03]

Peysakhov and Regli used assembly graphs to represent feature-based connectivity of LEGO assemblies. An assembly graph is very expressive and can represent a variety of LEGO assembly

structures comparing other representations. In assembly graphs, the nodes represent LEGO elements and the edges represent connections among the elements. They also proposed a graph grammar that can be used to evaluate the validity of the LEGO assembly structure [Pey03].

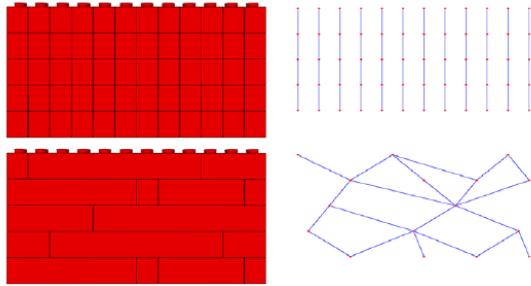


Figure 6. LEGO brick layouts (left) and corresponding graph representation [Tes13]

Testuz et al. proposed another graph representation for the LEGO structure. Fig. 6 shows the LEGO brick layouts and corresponding graph representations used to evaluate connectivity and stability of the construction. In their graph representation, a node denotes a LEGO brick and an edge denotes the connection between the bricks. The solidity optimization to improve the stability of the construction was conducted based on the assumption that the more LEGO bricks are connected, the stronger the connectivity will be [Tes13].

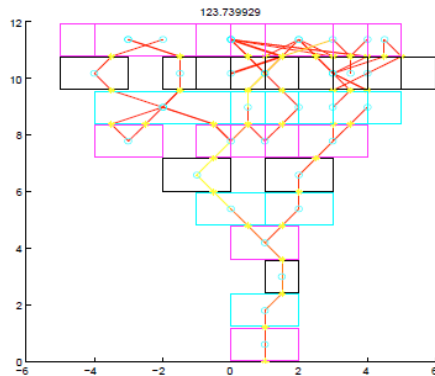


Figure 7 A brick layout and corresponding NTP representation in two-dimensional space [Fun98]

Funes and Pollack proposed a network of torque propagation structure to represent a LEGO assembly to evaluate the stability of the structure. A network of torque propagation (NTP) consists of: (1) a list of bodies, (2) a list of joints, (3) a list of forces, and a symbol G that denotes the "ground". Here a joint is defined as the center position of the area of contact between a pair of bricks [Fun01].

When using evolutionary algorithms to solve the problem, the solution itself of the problem must be encoded as genotype representations. There are two

approaches to represent genotype: one is direct and the other is indirect representation. Direct genotype representation is conceptually identical to the phenotype or the solution of the problem. In indirect representation phenotype can be constructed from the transformations of its genotype [Pey03].

The advantage of the indirect representation is that it can focus the search process through the feasible search space by significantly reducing the space. The disadvantage of the indirect representation however is that the standard genetic operators cannot be directly used [Pet01]

4. Cost Functions

The cost function for the optimization problem must be designed based on the performance criteria of the problem described in chapter 2. The most important factors to consider for cost function design are stability of the created LEGO assembly and the processing time to create it. Gower et al. introduced a set of heuristics that are useful in designing the cost function for the problem based on their rigorous research [Gow98][Sma08].

Gower et al. proposed six heuristics that are necessary to guarantee the stability of the created LEGO assembly. The first three heuristics are as follows: (1) A high percentage of the area of each brick should be covered by other bricks from above and below; (2) Larger bricks must be preferred over small bricks; (3) Bricks in consecutive layers should have alternating directionality [Gow98][Sma08].

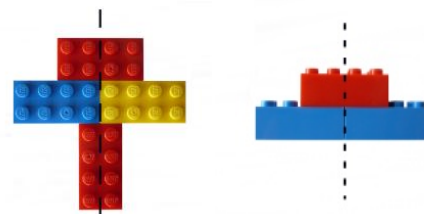


Figure 8 The boundary defined by the neighboring bricks (left) and the vertical boundary (right) [Sma08]

We need to be more careful at the boundaries of the whole model and vertical boundaries of each brick where connectivity is more vulnerable. The other three heuristics addressed the connectivity and stability problem at the boundaries: (1) A high percentage of the vertical boundaries of each brick should be covered by bricks in the consecutive layers; (2) A brick must be placed such that the middle of the side should be at the boundary defined by the neighboring bricks; (3) If a brick covers a vertical boundary in the previous layer, the middle of the brick must be aligned to the boundary [Gow98][Sma08].

Based on the heuristics Gower et al. defined a cost function as follows:

$$P = C_1P_1 + C_2P_2 + C_3P_3 + C_4P_4$$

where, P_1 relates to the alternating directionality, P_2 corresponds to coverage of the vertical boundaries, P_3 relates to the coverage of the boundary defined by the neighboring bricks, and P_4 encourages the use of larger bricks. C_i 's represent the weights for each term.

Peysakhov and Regli 03 proposed a more advanced and flexible form of the cost function to evaluate the ability of an LEGO assembly relative to its performance and function. Their cost function use the attributes of the LEGO structure including weight, number of nodes, and size of the structure. Their cost function is as follows:

$$\frac{1 + \sum P_i(a_i)}{1 + \sum P_i(b_i) + \sum P_i(|c_i - t_i|)}$$

Here, P_i is the weight function that represents the importance of the parameter. They set the weight value as the equal value to the parameter itself for the most important parameters. They set the weight value to the square root of the parameter for less important ones. For the least important parameters, they used square root of square root of the parameter.

$$P_i = x_i, P_i = x_i^{1/2}, P_i = x_i^{1/4}$$

a_i denotes the properties that will be maximized such as reliability. b_i denotes the properties that will be minimized such as manufacturing cost, and c_i denotes the properties that will be as close as to the specific constant value t_i [Pey03].

5. Solution Methods

A variety of approaches have been proposed to solve the LEGO construction problem. In this section we will describe and discuss those solution methods including greedy algorithms, local search, beam search, cellular automata, and evolutionary algorithms

Greedy Algorithms

Ono and Alexis 13 proposed a method to convert a 3D mesh model into a corresponding LEGO model by using their replacement strategy. The input to the system is the 3D model and user-specified level-of-detail. The system. The system converts the input 3D model into a voxel model based on the level-of-detail, and then converts it to the LEGO model [Ono13].

The system places the unit LEGO brick of the size 1x1 to each voxel. It then merges the unit bricks to replace each voxel with larger bricks so that the resulting LEGO structure would be more stable. They represent the LEGO structure as a legograph shown in chapter 3 with three different types of links they defined. The replacement is conducted layer-by-layer, from bottom to top and the replacement is performed in each layer using a greedy method. In

the replacement procedure, for each position the brick type with the highest score is chosen to be replaced. The strategy for the scoring is designed to guarantee the stability of the resulting LEGO structure and it is similar to the cost function described in chapter 4. When the LEGO structure is built, their system automatically generates the assembly instructions [Ono13].

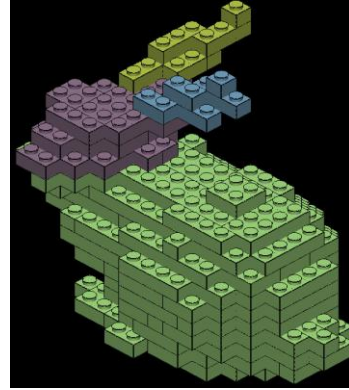


Figure 9. The result of greedy method by [Ono13]

Testuz et al. proposed a similar method to [Ono13] in that they fill the unit bricks into each voxel first and then merge and split the bricks sequentially to obtain the optimal layout using greedy method. Their merge algorithm randomly select a brick and find a legal set of neighbors. It then repeat choosing the neighbor with the lowest cost and select the neighbor with the lowest cost value until there are no more mergeable neighbors. This process repeats until there is no more brick to merge [Tes13].

In building the LEGO model Testuz et al. evaluate the stability of the model as other approaches do. To achieve this, they used the graph representation described in chapter 3. In the stability evaluation, they assumed that the stability will be stronger if the more bricks are connected to each other [Tes13].

Local search

In each step of the procedure, local search approach considers only a small subregion and attempt to find the best brick placement to fill the subregion [Gow98][Sma08]. Only a few bricks are permanently placed in each step considering the effect of the local placement for the global solution. Then the subregion slightly moves so that a new subregion overlaps the previous one as a sliding window [Sma08].

In this approach, the important issue is the size of the subregion. If the size of the subregion is too small, it is hard for the local placement contribute to the global optimization. On the other hand, if the size of the subregion is too large, the search space would be larger resulting in increase in processing time. The optimal size for the subregion therefore must be determined based on the size and characteristics of the input real-world object. [Sma08]

To apply simulated annealing to the LEGO construction problem, we can firstly divide each layer into subregions of smaller size. Then the subregions will be randomly filled with arbitrary placements of LEGO bricks resulting in the initial state. For each subregion a set of successor states are generated by replacing a small number of bricks with new bricks. New successor states are generated until we find a new state that decreases the energy. The search process will stop after the number of iterations specified by a user is complete or when an acceptable solution is found [Sma08].

Simulated Annealing

Simulated annealing is a variant of the hill-climbing technique that computes all possible successor states from the current state and then selects the best successor. The well-know limitation of the approach is that it can easily converges to the local minimum instead of the global optimum [Sma08].

At each iteration, simulated annealing algorithm considers a set of neighboring states from the current state. It probabilistically compares between moving to new states and staying in the current state and then decide the new state that minimizes the energy. This process repeats until it finds a satisfactory solution [Sma08].

Beam Search

A beam search is conceptually similar to the simulated annealing approach in which successor states are generated and evaluated to find a new state with better quality. A beam search approach a best-first search algorithm and it is different from the simulated annealing approach in that all the possible successor states are generated and evaluated using a cost function to find the new state with the best cost. The algorithm therefore searches for the best local solution at each step [Sma08].

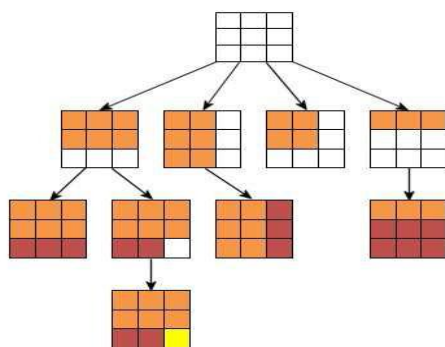


Figure 10. Beam search tree to a fill 3x3 layer using the standard LEGO bricks [Sma08]

At each step a beam search algorithm finds best k successors and they are added to their parent states. Then the search process continues while pruning the

states that cannot generate any successor states of better quality from the search tree. The problem however is that it can focus on a too narrow search space resulting in not convincing solutions. An improvement for this problem is to select the k successors probabilistically with a higher probability of selecting the lower cost successors to create a broader search space [Sma08].

Cellular Automata

van Zijl and Smal proposed an approach using cellular automata based on the cost function proposed by [Gow98][Van08]. Their approach is conceptually similar to the merge/split approach using several heuristics that guides the search [Tes13]. The approach virtually cuts the given 3D object into horizontal two-dimensional layers. It finds the optimal 2D layout first and then join them to construct final 3D model. If we solve a 2D layout optimization problem separately, the stability of the resulting model cannot be guaranteed. They therefore used the Gower et al.'s heuristics during each step to solve the 2D problem to guarantee the solidity of the model [Gow98][Van08].

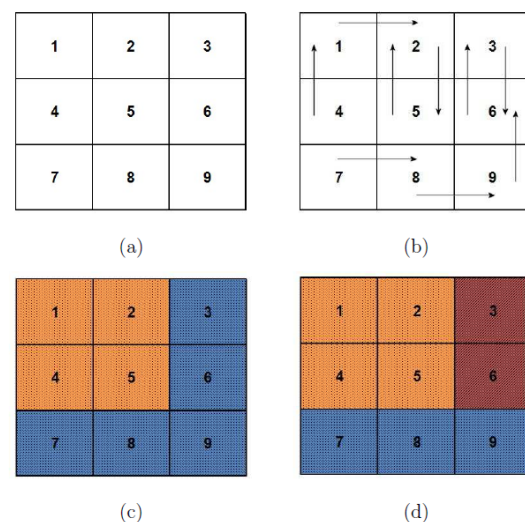


Figure 11. An example of cellular automata representation (a) the 2D grid, (b) potential merge neighbors, (c) potential new clusters, (d) the final three clusters [Sma08]

This approach used a legolised representation described in chapter 3. In initial stage, each cluster of unit size 1x1 that contains value one represent a unit LEGO brick. For each cluster the algorithm checks if it can be merged with any of its Von Neumann neighborhood. Two clusters can be merged if the merge can result in a new cluster that represent a larger valid LEGO brick. This merge operation is conducted for all the clusters in the layout. The order of merges can be random, front-to-back, or any other orders chosen by a user [Van08].

Evolutionary Algorithms

Evolutionary algorithms are very effective optimization technique for the problems whose optimal solution is hard to formalize. LEGO construction problem is a hard combinatorial optimization problems in which it is infeasible to find the optimal solution and the "good" solutions of reasonable quality are enough. Evolutionary algorithms could be a proper approach to solve the problems that have such features and nature [Pey03][Pet01].

To solve an optimization problem using evolutionary algorithms, we have to encode the solution of the problem as chromosomes, define the evaluation function, and develop mutation and recombination operators depending on the characteristics of the given problem. We have discussed about the genotype representation for LEGO construction problems in chapter 3 and we will therefore discuss about evaluation functions and operators developed to solve the problem using evolutionary algorithms [Pey03][Pet01].

There are two approaches to genotype representations: one is direct representation, and the other is indirect representation. In direct representation the genotype is conceptually identical to the corresponding phenotype. On the other hand, in indirect representation, the genotype is transformed to construct the corresponding phenotype. Indirect representation usually have more information about the phenotype and it therefore can focus the search space by reducing the space. The problem of indirect representation is that the standard operators such as mutation and crossover do not directly work. We therefore have to redefine the operators according to the structure of genotype [Pey03][Pet01].

0	Plate(6,2)	0>2	Snap[(1,2)(1,1)]
1	Plate(6,2)	0>3	Snap[(6,2)(2,1)]
2	Brick(2,4)	1>2	Snap[(1,1)(1,4)]
3	Brick(2,4)	1>3	Snap[(6,1)(2,4)]
4	Beam(4,1)	2>4	Snap[(2,2)(1,1)]
		3>4	Snap[(1,2)(1,4)]

Figure 12. An example of genotype representation by [Pes03]

Peyssakhov and Regli developed their chromosomes using a combination of two data structures: one is an array of all nodes, and the other is the adjacency hash table containing all edges as shown in fig 12. The key value of the hash table represents the position and direction of edges. For example, the key "1>3" means that the edge connects from the node 1 to the node 3. Hash table also describe how the LEGO elements are connected [Pey03].

The mutation operator of [Pey03] is applied with constant and low probability to provide the balance between the exploration and exploitation. When a mutation arises for a node, a LEGO brick is simply replaced by the same type brick with different size [Pey03].

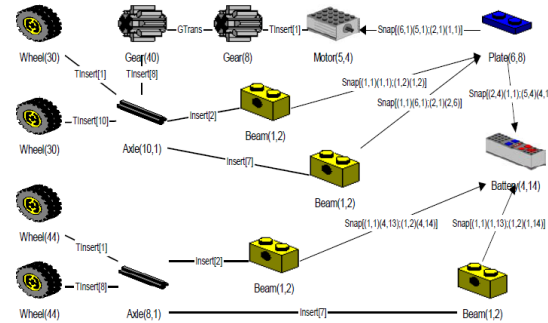


Figure 13. An example assembly graph for the LEGO car [Pey03]

Crossover is conducted by two operators: cut and splice. It selects two chromosomes for crossover and random points are selected respectively for the two chromosomes by cut operator. The tail parts of the parent chromosomes are then spliced with the head parts of them [Pey03].

Petrovic proposed more advanced and complicated operators as follows. His crossover operator firstly selects a rectangular region at random. Then a part of LEGO bricks are copied from one parent and other bricks that do not conflict with already placed bricks are copied from another parent [Pet01].

Petrovic suggested the following mutation operators because random mutation operator can generate overlaps.

- A brick is replaced by other random brick.
- A brick is added to an empty location randomly.
- A brick is shifted by one unit in one of the four possible directions.
- A brick is eliminated from the layout
- A brick is extended by one unit in one of the four possible directions.
- All bricks that are in a random rectangle are replaced by random bricks
- The whole layout is initialized again

In his mutation operators, larger bricks are always preferred to be replaced to increase stability of the structure [Pet01].

6. CONCLUSIONS

In this paper, we reviewed a variety of research efforts to address the automated LEGO construction problem. We investigated the problem definition and formulation, various data representations for 3D polygonal mesh models and LEGO assembly structures, cost functions to solve the optimization problem for LEGO construction and solution

methods a number of researchers proposed. To date, graph representations have been widely used to represent the LEGO structure and as solution methods, greedy algorithms, simulated annealing, beam search, cellular automata, and evolutionary algorithms have been used to automatically construct LEGO structure minimizing the number of bricks used and guaranteeing the stability of the built structure. Those approaches are useful to create a LEGO structure design for given 3D polygonal models for entertainment purposes and also can be useful for engineering education.

7. REFERENCES

- [Cam12] Campbell D., Freidinger E., Querns M., Swanson S., Ellis A., Kuech T., Payne A., Socie B., Condren S. M., Lisensky G., Rasmussen R., Hollis T., Villarreal R., Campbell K., Exploring the Nanoworld with LEGO Bricks, A Free Book, Materials Research Science and Engineering Center, University of Wisconsin-Madison, 2012. <http://education.mrsec.wisc.edu/LEGO/PDFfiles/nanobook.PDF>
- [Fun01] Funes P. J., Pollack J. B. Componential Structural Simulator. Technical Report, Brandeis University, 1998.
- [Gow98] Gower R., Heydtmann A., Petersen H. LEGO: Automated Model Construction. Jens Gravesen and Poul Hjorth, pp. 81-94, 1998.
- [Iga11] Igarashi Y., Suzuki H. Cover geometry design using convex hulls. Computer-Aided Design, 43, 9, pp. 1154-1162, 2011.
- [Lam06] Lambrecht, B. Voxelization of boundary representations using oriented LEGO plates. University of California, Berkeley, 2006. http://lego.bl.design.org/LSculpt/lambrecht_legovoxels.pdf
- [Lo09] Lo K. Y., Fu C. W., Li H. 3D polyomino puzzle. ACM Transactions on Graphics, 28, 5, Article No. 157, 2009.
- [Mit04] Mitani J., Suzuki H. Making papercraft toys from meshes using strip-based approximate unfolding. ACM Transactions on Graphics, 23, 3, pp. 259-263, 2004.
- [Ono13] Ono S., Alexis A. Automatic generation of LEGO from the polygonal data. International Workshop on Advanced Image Technology, pp. 262-267, 2013.
- [Pet01] Petrovic P. Solving the LEGO brick layout problem using evolutionary algorithms. Technical Report, Norwegian University of Science and Technology, 2001.
- [Pey03] Peysakhov M., Regli W. Using Assembly Representations to Enable Evolutionary Design of Lego Structures. Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 17:155-68, 2003.
- [Sil09] Silva L., Pamplona V., Comba J. Legolizer: A real time system for modeling and rendering LEGO representations of boundary models. XXII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI), pp. 17-23, 2009.
- [Sma08] Smal E. Automated Brick Sculpture Construction. MS. Thesis, The University of Stellenbosch, 2008.
- [Tes13] Testuz R., Schwartzburg Y., Pauly M. Automatic generation of constructable brick sculptures. Eurographics 2013 Short Papers, pp. 81-84, 2013.
- [Tim98] Timcenko O. LEGO: How to build with LEGO. 32nd European Study Group with Industry Final Report, pp. 81-94, 1998. <http://www.maths-in-industry.org/past/ESGI/32/Report/ESGI32.ps>
- [Van08] van Zijl L., Smal E. Cellular automata with cell clustering. Proceedings of Automata 2008 Workshop, Bristol, UK, pp. 425-440, 2008.
- [Wan08] Wang W., Lu A., Yu Li., Li Z. A digital lego set and exercises for teaching security protocols. the 12th Colloquium for Information Systems Security Education, University of Texas, Dallas, TX, June 2-4, 2008.
- [Win05] Winkler D. Automated brick layout. BrickFest, 2005. <http://www.brickshelf.com/gallery/happyfrosh/BrickFest2005/automatedbricklayout.pdf>
- [Xin11] Xin S., Lai C. F., Fu C. W., Wong T.T., He Y., Cohen-Or D. Making Burr Puzzles from 3D models. ACM Transactions on Graphics, 30, 4, Article No. 97, 2011.
- [Yip11] Yip-Hoi D. M., Newcomer J. L. Teaching CAD modeling using LEGO. American Society for Engineering Education, 2011. <http://www.asee.org/public/conferences/1/papers/152/download>

A Control Cluster Approach to Non-linear Deformation

A. A. Bukatov
SFU Computer Center
Bolshaya Sadovaya 105, 344000,
Rostov-on-Don, Russia
baa@sfedu.ru

E. E. Gridchina
Informatics and Computational
Experiment Department, SFU,
BolshayaSadovaya 105, 344000,
Rostov-on-Don, Russia
helen.gridchina@gmail.ru

D. A. Zastavnoy
Informatics and Computational
Experiment Department, SFU,
BolshayaSadovaya 105, 344000,
Rostov-on-Don, Russia
dzast@sfedu.ru

ABSTRACT

Modeling plausible deformation of the objects has been an important task in computer animation and game design industry. The approach proposed in the paper deals with a polygonal mesh deformation splitting the vertices of the mesh into two types: cluster vertices and free vertices. With the user defining the shape of the mesh key areas with the help of cluster vertices, the algorithm takes advantage of non-linear geometric deformation for calculating free vertices position. The approach could be used both for creating a sequence of altered model shapes to produce a character animation (with the help of user-created control cluster data) and for visualizing some ecological processes.

Keywords

Control cluster, non-linear deformation, skeletal animation, polygonal mesh deformation.

1. INTRODUCTION

Deforming model meshes has become an active field of research [1-5]. It can be used both for modeling characters and for creating a sequence of altered model shapes to produce an animation.

Most of the algorithms can present either physical or geometric approach. We concentrate in the geometric approach domain. Most of the geometric algorithms proposed demonstrate shortcomings connected with their linear approach. So this research focuses on the non-linear geometric approach to deformation.

In this paper a geometric method for plausible polygonal model deformation is proposed. It makes use of control clusters to facilitate the work of the animator in creating realistic animation of the character without involving much user-input data. The method proposed provides intuitive control and it is easy to use because it allows the user to influence only a small group of vertices leaving the non-linear deformation of the rest of the vertices to the algorithm.

In the previous papers an approach based on a control clusters technique for correcting 2D models skinning

deformation was presented [6]. This paper is to describe a generalized and more versatile approach to deforming 3D polygonal models – Control Cluster Method (CCM). The generalized approach can be implemented in a wide variety of applications including physical processes simulation, such as ecological processes (e.g., computational domain flooding and dewatering as a result of wind upsurge-downsurge; fire spreading; prevalence of air and water pollution).

2. RELATED WORK

Among the geometric deformation technologies connected with skeletal animation the most widely used is Linear Blend Skinning (LBS). Being simple and straightforward LBS is notorious for its well-studied shortcomings [7].

To avoid well-known artifacts Pose Space Deformation proposed in [8] takes advantage of using sample shapes of the model, the technique requiring much input from the user [9].

Dual-quaternion skinning is a more recent skeletal animation technique [10], describing both rotations and translations using quaternions. Having no shortcomings of LBS, dual-quaternion skinning demonstrates some new artifacts connected with too much volume.

It was also proposed in [11] to pre-compute optimized skinning weights for linear and dual-quaternion skinning at joints to approximate the skin transformations produced by nonlinear variational deformation methods.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

[12] exploits the advanced compositions mechanisms of volumetric implicit representations for correcting the results of geometric skinning techniques.

3. PROPOSED APPROACH

3.1 General description

Control cluster method assumes all the vertices of the model are split into two types: cluster vertices and so-called free vertices. The position of the cluster vertices is defined by some input data (see 3.4). The key areas of the mesh defined by cluster vertices form the shape of the model, thus cluster vertices are used for controlling the deformation of the model. The position of free vertices is calculated automatically.

3.2 Cluster vertices transformation

Cluster vertices can be transformed in a great variety of ways, either by mesh editing or by applying some deformer including lattices, cage or skeleton. Cluster vertices transformation could be also defined by data from some type of measuring equipment, for example water level measuring device.

3.3 Free vertices transformation

If a vertex does not belong to the cluster, it is considered a free one and its position is recalculated based on its position in the model topology and its nearest cluster vertices position. Free vertices position can be defined in several non-linear ways, including cubic cardinal splines generalizing cubic Catmull-Rom splines [13] so that it corresponds to the position of the cluster vertices.

Using interpolation, C^1 -continuity and local control, cardinal splines are an acceptable way to solve the task. The C^1 -continuity of the spline provides a smooth natural look of the character, with the spline remaining flexible. Also the spline interpolates its control points giving direct control over the points of the curve. With local control the spline has every control vertex provide a slight impact on the overall look, so the model details are preserved. Therefore cardinal splines allow the achievement of realistic deformation of commonly difficult parts of the model.

The shape of the curve the spline gives depends heavily on the parameterization defined [14]. Choosing a spline parameterization for our method we considered three types: uniform parameterization, chord-length parameterization, centripetal parameterization.

Uniform parameterization is considered the most popular choice for cardinal splines, though for curves with segments of different length this parameterization often leads to artifacts such as self-intersections within short curve segments [14], which is as usual unacceptable. Cusps and intersections are also possible when using chord-length

parameterization, the curve “overshoots” within longer curve segments. Centripetal parameterization is the only not to generate such artifacts. Moreover, among these parameterizations the centripetal version appears to produce a curve that is closer to the control polygon than the others.

Despite the fact that it is mathematically proven that centripetal parameterization lacks traditionally undesired features such as cusps and intersections within a segment [14], the CCM uses chord-length parameterization. Our reasoning behind this preference differs from the standard one as CCM can deal with character models. Firstly, the curvature of the relatively long curve segments is larger in comparison with the results of the other parameterizations considered, and it tends to remain small within shorter curve segments. It helps achieve a more realistic look of the character, as human-like models will not lose much volume when animated. Secondly, chord-length parameterization is considered the best as it provides a very well-conditioned linear system of equations compared to other parameterization types. So if p_i, p_{i+1} are spline control points the parameterization $t_{i+1} = t_i + |p_{i+1} - p_i|$ is used.

The cardinal spline curve segment shape depends on four neighboring data points: $p_{i-2}, p_{i-1}, p_{i+1}, p_{i+2}$. The tangent is calculated as

$$m_k = (1 - q) \frac{p_{k+1} - p_{k-1}}{t_{k+1} - t_{k-1}} \quad (1)$$

The tension parameter q defines the curvature of the spline, $q \in [-1, 1]$. The method sets $q = -0.5$ as the default parameter value.

3.4 Fields of application

Control cluster method could be used in different spheres of visualization ranging from animation to ecological processes simulation.

The cluster vertices deformation can be set using user-input data, including different deformer such as a skeleton to produce a character animation or some certain pose of the model.

Provided cluster vertices positions are defined by some scientific equipment, a wind upsurge-downsurge flooding and dewatering simulation could be achieved with the help of control cluster method. Among possible spheres of control cluster method application in ecological processes simulation fire spreading and prevalence of air and water pollution can be mentioned.

In this paper application of the method to animation of human-like character models is presented.

4. APPLICATION TO SKINNING

4.1 Skinning cluster vertices

transformation

Skinning application of CCM can be used for creating character animation. Generally speaking, cluster vertices transformations can be defined in different ways that are not strictly set. To demonstrate it we consider three different ways for cluster vertices transformation in creating arm rotation animation: Linear Blend Skinning, Pose Space Deformation and procedural dependence.

The most efficient, versatile and wide-spread way of creating a character animation is Linear Blend Skinning. So base cluster vertices can be deformed with LBS.

Let $B=\{b_i\}$ be a skeleton, i.e. a hierarchy of bones. Every bone b_i is assigned a coordinate system and a 3D transformation (translation, rotation and scaling), defined by matrix W_i . The transformation of a child node of the hierarchy inherits its parent node

transformation. Every vertex $v \in V_P$ is associated with a set of weights $\{w_i\}$, $\sum w_i = 1$, where w_i is the weight of the bone b_i . Weight w_i defines the extent to which the vertex position is influenced by the bone b_i . Let $\{B_i\}$ be the skeleton configuration in the bind pose. The skeleton being in an arbitrary pose $\{W_i\}$, the transformed position of cluster vertex v' is calculated according to the formulae

$$v' = LBS(v) = \sum_i w_i W_i B_i^{-1} v \quad (2)$$

Being versatile and computationally efficient, LBS demonstrates some undesired artifacts, such as volume loss. One of the defects, so-called “collapsing elbow”, results in unnatural look of the character. To correct those undesired artifacts PSD can be used. It requires using sample pairs $\langle X, S \rangle$, where X is a user-input sample shape of the model corresponding to skeleton configuration S . With the flexibility of control cluster method it is possible to avoid defining sample shapes of the whole model. Instead only the problem vertices of the elbow area of the model can be defined as cluster ones and only one sample pair is used for deforming those vertices likewise PSD thus achieving the necessary amount of volume. So cluster vertices of the elbow problem area undergo additional transformations with the help of the displacements

Let us assume that v is a vertex position in the sample pose of cluster vertices in the pose X^i , v^0 is vertex position in the bind pose B . Then displacement d_j in the bone b_j local coordinate system is calculated as follows:

$$d_j = LBS^{-1}(v, b_j) - B_j^{-1} v^0 \quad (3)$$

If the current pose $X=X^i$, the current position of the control cluster vertex v is calculated as:

$$v' = LBS(v, d) = \sum_j w_j W_j (B_j^{-1} v + d_j) \quad (4)$$

Another welcome feature for realistic animation is muscle bulging. It could also be achieved using PSD but it would require for the user to create some sample pairs to imitate realistic muscle bulging. As an alternative way cluster vertices corresponding to biceps area of the upper arm can be deformed procedurally depending on the angle between upper arm and lower arm bones.

5. RESULTS

For the sake of example cylinder model (a simplified “hand” model) with rigging is used (see Fig. 1). All the vertices are split into 4 groups: 3 cluster groups and free vertices.

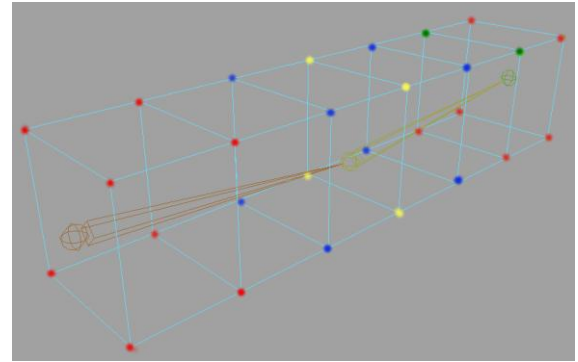


Figure 1. The model of a cylinder with a two bone skeleton. Red – cluster LBS deformed vertices; yellow – cluster likewise PSD deformed vertices; green – cluster procedurally deformed vertices; blue – free vertices.

The base cluster vertices of the model are deformed with help of LBS, they define the overall shape of the model. Cluster vertices of the “elbow” problem area of the model are transformed using one sample pair for the four cluster vertices, the sample pair being extracted from PSD sample shape of the whole model. To create biceps bulging effect corresponding cluster vertices position is calculated procedurally based on the angle between the two bones.

Comparing the three methods (LBS, PSD and CCM) in a rotation animation without muscle bulging it is possible to say that yielding close results neither PSD, nor CCM demonstrates the volume loss common for LBS (see Fig. 2). As far as user input data is concerned, PSD requires to create at least one 12 vertices sample pair for the animation, while it is enough to store one 4 vertices sample for CCM.

Moreover, with a greater angle of rotation due to using cardinal splines CCM forms a cusp in the area

of the elbow that is natural for human like characters (see Fig. 3).

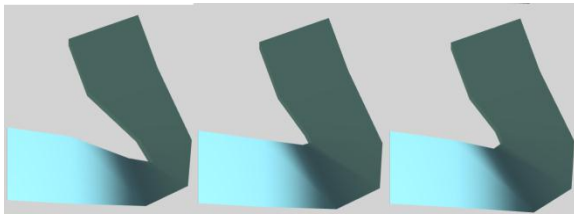


Figure 2. Side view of the model, frames of the child bone rotation animation. From left to right: LBS, PSD, CCM.

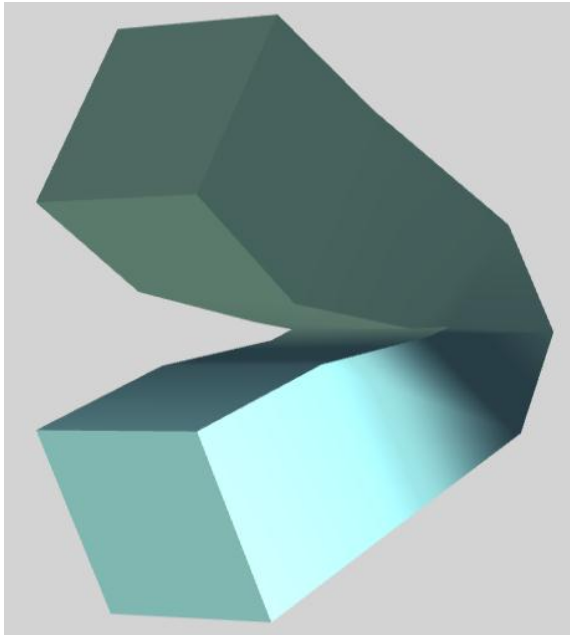


Figure 3. CCM rotation animation of the model with a slight muscle bulging effect.

6. CONCLUSION AND FUTURE WORK

Control cluster method is a novel approach for geometric non-linear deformation of the model in a labor-saving way. In this paper its application to skinning is described. Applications of the approach to other fields as physical processes simulation is in the scope of the future research.

7. REFERENCES

- [1] Gain J., Bechmann D., A survey of spatial deformation from a user-centered perspective. *ACM Trans. on Graphics*, 27/4, 107:1–107:21 (2008).
- [2] Cohen-Or D. Space deformations, surface deformations and the opportunities in-between. *J. Comput. Sci. Tech.*, 2009,24: pp. 2–5.
- [3] Botsch M, Sorkine O. On linear variational surface deformation methods. *IEEE Trans. Vis. Comput. Graph.*, 2008, 14: pp. 213–230.
- [4] Deformation Models: Tracking, Animation and Applications Series: Lecture Notes in Computational Vision and Biomechanics, Vol. 7, González Hidalgo, Manuel, Mir Torres, Arnau, Varona Gómez, Javier (Eds.), 2013, XIV, 297 p.
- [5] Botsch M., Kobbelt L., Pauly M., Alliez P. and Levy B., *Polygon Mesh Processing*. AK Peters - 2010.
- [6] Bukatov A.A., Gridchina Y.Y., Zastavnoy D.A. A Spline-Based Approach to Control Cluster Deformation, *World Applied Sciences Journal*, 26 (6): 724-727, 2013.
- [7] A. A. Bukatov, E. E. Gridchina, D. A. Zastavnoy, 2012. Skeletal animation techniques for deforming polygonal surfaces of 3D models. *Inzhenerniy Vestnik Dona*, 3: 59-74 (in Russian).
- [8] J. P. Lewis, M. Cordner and N. Fong, 2000. Pose space deformation: a unified approach to shape interpolation and skeleton driven deformation. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, pp: 165-172.
- [9] G. S. Lee and F. Hanner, 2009. Practical experiences with pose space deformation. In *ACM SIGGRAPH ASIA 2009 Sketches*, pp: 43:1-43:1.
- [10] Skinning with dual quaternions / L. Kavan, S. Collins, J. Zara, C. O'Sullivan // *Proceedings of the 2007 symposium on Interactive 3D graphics and games*.— I3D '07.— New York, NY, USA: ACM, 2007.— Pp. 39–46.
- [11] Kavan, L. Elasticity-Inspired Deformers for Character Articulation / L. Kavan, O. Snorkie // *ACM Trans. Graph.* 31(6): 196 (2012).
- [12] Implicit skinning: real-time skin deformation with contact modeling / Vaillant R., Barthe L., Guennebaud G., Cani M., Rohmer D., Wyvill B., Gourmel O. and Paulin M. // *ACM Trans. Graph.* 32, 4, Article 125 (July 2013), 12 pages.
- [13] E. Catmull and R. Rom, 1974. A class of local interpolating splines. In *Computer Aided Geometric Design*, Eds., R. E. Barnhill and R. F. Reisenfeld. New York: Academic Press, pp: 317-326.
- [14] C. Yuksel, S. Schaefer and J. Keyser, “On the parameterization of Catmull-Rom curves”, *ACM Joint Conference on Geometric and Physical Modeling (SPM '09)*. New York: ACM, 2009, pp. 47–53.